Modeling and Analysis of Nonlinear Control Systems Using Exterior Differential Systems

by

David Donald Niemann

B.S. (Kansas State University) 1986
M.S. (Kansas State University) 1988

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor J. Karl Hedrick, Chair
Professor Shankar Sastry
Professor Andrew Packard

1996

The dissertation of David Donald Niemann is approved by:

Chair Date

Date

Date

University of California at Berkeley

1996

Modeling and Analysis of Nonlinear Control Systems Using Exterior Differential Systems

Abstract

Analysis and Design of Nonlinear Control Systems Using Exterior Differential Systems
by
David Donald Niemann
Doctor of Philosophy in Mechanical Engineering
University of California at Berkeley
Professor J. Karl Hedrick, Chair

This dissertation discusses the modeling and analysis of nonlinear control systems using exterior differential systems. The objectives of the dissertation are to provide a self-contained treatment of the theory of exterior differential systems which focuses on the aspects of the theory most applicable to the study of nonlinear control systems, to show how exterior differential systems theory can be applied to the study of nonlinear control problems, and to explore the relationship between the established vector field approach to geometric nonlinear control theory and the exterior differential systems approach.

The dissertation is divided into three parts. Part I contains the introductory material. Part II develops the background material on exterior differential systems which is needed to study nonlinear control systems. The presentation is arranged so that the topics which only involve finite-dimensional vector spaces are discussed first. This material includes tensors, forms, Grassmann manifolds, and systems of exterior equations. Next, the concept of a fibre bundle is introduced, and the discussion shifts to techniques which can be used to extend the structures defined over a finite-dimensional vector space to fields defined over the tangent bundle of a differentiable manifold. This material includes a discussion of vector fields, tensor fields, Grassmann bundles, distributions, codistributions, and exterior differential systems.

Part III uses the theory developed in Part II to model nonlinear control systems. This material begins with a presentation on how a Grassmann bundle can be used to model an affine nonlinear control system and a discussion of a prolongation process which can be used to model dynamic state feedback. The last chapters of this section provide a comparison with the geometric theory based on smooth vector fields. The topics discussed include invariant distributions, controlled invariant distributions, controllability distributions, the disturbance decoupling problem, and the noninteracting control problem.

Based on these results, it appears that much of the existing theory can be reinterpreted using exterior differential systems and that in many cases the results obtained using exterior differential systems represent a generalization of the standard results in the vector field approach.

Dissertation Chair

Contents

List of Figures

Part I

Preliminaries

Chapter 1

Introduction

Recently, there has been interest in applying a body of mathematics known as exterior differential systems theory to nonlinear control problems. This work originated around 1992 as a new method of studying the geometric properties of control systems such as controllability and feedback linearizability. Gardner and Shadwick [1] wrote an influential paper in which they formulated the conditions for feedback linearizability in terms of exterior differential equations. Significant contributions have also been made by Murray [14], Sluis [20], and Tilbury [18, 12].

Murray and Sastry [2] provided the first applications of this new approach in connection with their study of nonholonomic path planning. Most of their work has centered on path planning for steered vehicles such as truck-trailer combinations. These systems are characterized by the fact that the velocity vector is subject to nonintegrable constraint equations which arise from the no-slip conditions at the tires. Such constraints are called nonholonomic. The presence of nonholonomic constraints makes these systems difficult to analyze using the traditional vector field approach. But viewed as an exterior differential system, the nonholonomic constraints can be incorporated into the system in a very natural way. Using this approach, Tilbury et. al. [3] have been able to generate steering algorithms for very general multi-trailer combinations. From these results, it appears that the exterior differential systems approach to studying nonlinear control problems may offer some enhancements to the more established vector field approach.

The purpose of this dissertation is to further develop techniques for modeling and analyzing nonlinear control systems using exterior differential systems. More specifically, the dissertation has three main objectives: to provide a self-contained treatment of the theory of exterior differential systems which focuses on the aspects of the theory most applicable to the study of nonlinear control systems, to show how exterior differential systems theory can be applied to the study of nonlinear control problems, and to explore the relationship between the established vector field approach to geometric nonlinear control theory and the exterior differential systems approach.

The first objective is to provide a self-contained treatment of the theory of exterior differential systems which focuses on the aspects of the theory most applicable to the study of nonlinear control systems. Much of the existing mathematical literature on exterior differential systems is written at a very high level and is not oriented towards applications in control theory. More elementary treatments of some topics do exist, but they do not cover all the necessary material.

The second objective is to show how a nonlinear control system can be viewed as a geometric object using the exterior differential systems approach. This object is called a Grassmann bundle, and its structure incorporates both state transformations and affine state feedback. Although the Grassmann bundle is an abstract object, it is a useful tool for visualizing a control system in a coordinate-free way, and it often provides nice “pictorial” representations of theorems from nonlinear control theory. The Grassmann bundle model of a control system can also be used to provide a geometric model of a dynamic compensator. To some extent, this model allows one to visualize the effect that the dynamic feedback will have on the nonlinear system.

The third objective is to explore the relationship between the established vector field approach to geometric nonlinear control theory and the exterior differential systems approach. To this end, the dissertation discusses how standard topics from the vector field theory such as invariant distributions, controlled invariant distributions, and controllability distributions can be interpreted within the exterior differential systems framework. In fact, in many cases the results obtained using the exterior differential systems approach are actually more general because time variations in the system parameters or inputs can be handled with minimal complication.

In order to facilitate this presentation, the dissertation is divided into three parts. Part I contains the introductory material. Part II develops the background material on exterior differential systems which is needed to study nonlinear control systems. Unfortunately, there is quite a bit of machinery, but the sections are fairly self-contained, and some parts can probably be skipped depending on the reader's familiarity with the various topics. Part III uses the theory developed in Part II to model nonlinear control systems. The emphasis is on comparisons with the geometric theory based on smooth vector fields defined over a manifold. This approach is typified in the books by Isidori [4] and Nijmeijer and van der Schaft [23]. We will see that in many ways, the application of exterior differential systems to control theory represents a direct extension of this established geometric approach.

Chapter 2

Modeling Systems on the State-Time Space

Most of the basic elements in the exterior differential systems approach to modeling a nonlinear control system can be introduced without using all the mathematical machinery which will be developed in Part II of this dissertation. Therefore, in this chapter we will develop some basic ideas about modeling differential equations and nonlinear control systems on the state-time space. Although the mathematical prerequisites have been kept to a minimum, it is assumed that the reader has a basic understanding of the elements of differential geometry. If needed, additional background material on this topic can be found in Appendix A.

The material in this chapter is divided into two sections. The first section discusses geometric interpretations of a differential equation and its solutions. Particular emphasis is placed on the idea that we can view the graph of a solution to a differential equation as a one-dimensional submanifold of the state-time space. The second section generalizes the discussion to include affine control systems, and concludes by asserting that affine control systems can be represented as a fibre bundle over the state-time space. Hopefully, this presentation will help to motivate the mathematical material contained in Part II and to give the reader some idea where we are eventually headed.

2.1 Modeling Differential Equations on the State-Time Space

We will begin by considering a differential equation defined over $\mathcal{R}^n$ by a smooth vector field

$$\dot{x} = f(x). \tag{2.1}$$

Geometrically, we can picture the vector field as an assignment of one vector to each point of $\mathcal{R}^n$. A solution to the differential equation through $x_0 \in \mathcal{R}^n$ is a function $s_f : (-\epsilon, \epsilon) \to \mathcal{R}^n$, which satisfies $s_f(0) = x_0$ and $\dot{s}_f(t) = f(s_f(t))$ for every $t \in (-\epsilon, \epsilon)$. Any such function is called an integral curve of the vector field f. The image of this curve forms a one-dimensional submanifold of $\mathcal{R}^n$ whose tangent space at each point $s_f(t)$ contains the vector $\dot{s}_f(t)$. The theory of ordinary differential equations guarantees that there is a unique integral curve of f passing through each point $x_0 \in \mathcal{R}^n$. Depending on the vector field, the parameter ϵ may be limited, or we may be able to expand the interval $(-\epsilon, \epsilon)$ to be the whole real line. Since the function s_f is unique, it has a unique graph

$$(s_f(t), t) \subset \mathcal{R}^n \times (-\epsilon, \epsilon).$$

In the case $n = 1$, this corresponds to the usual notion of the graph of a function $y : \mathcal{R} \to \mathcal{R}$. Figure 2.1 gives a pictorial representation of these concepts. The picture on left shows the integral curve of the differential equations

$$\begin{aligned}\dot{x}^1 &= -x^1 - 2x^2 \\ \dot{x}^2 &= 8x^1 - .5x^2\end{aligned}$$

which passes through the point $x^1 = 1$, $x^2 = 1$ at time $t = 0$. The picture on the right illustrates the graph of this curve in the state-time space.

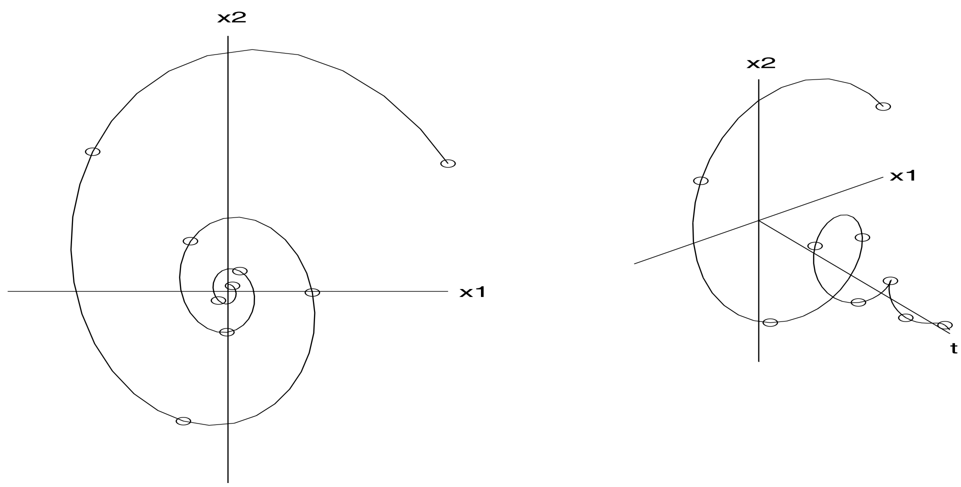

Figure 2.1: An Integral Curve and Its Graph

We can construct the graph of the integral curve s_f by taking the image of the integral curve of the extended vector field

$$\begin{aligned}\dot{x} &= f(x) \\ \dot{t} &= 1\end{aligned} \tag{2.2}$$

which passes through $(x_0, 0) \in \mathcal{R}^n \times \mathcal{R}$. As before, the theory of ordinary differential equations ensures that there is a unique integral curve, $\hat{s}_f : (-\epsilon, \epsilon) \rightarrow M \times \mathcal{R}$ which satisfies equation 2.2 and the initial condition $\hat{s}_f(0) = (x_0, 0)$. Furthermore, it is not hard to see that the image of this curve corresponds with the graph of s_f. However, it is important to realize that this is not the only vector field which generates an integral curve whose image corresponds with the graph of s_f. Any vector field which is a pointwise scalar multiple of the extended vector field 2.2 will also satisfy these conditions.

Example: Consider the smooth vector field

$$\begin{aligned}\dot{x} &= 5x \\ \dot{y} &= 3x + 4y\end{aligned} \tag{2.3}$$

which is defined over $\mathcal{R}^2$ with respect to the coordinates (x, y). The integral curve of this vector field which passes through the point (x_0, y_0) when $t = 0$ is

$$\begin{aligned}s_x(t) &= e^{5t}x_0 \\ s_y(t) &= 3(e^{5t} - e^{4t})x_0 + e^{4t}y_0,\end{aligned} \tag{2.4}$$

and its graph is described by

$$(s_x(t), s_y(t), t) \subset \mathcal{R}^2 \times \mathcal{R} \tag{2.5}$$

Equation 2.3 can be extended to the vector field

$$\begin{aligned} \frac{dx}{d\tau} &= 5x \\ \frac{dy}{d\tau} &= 3x + 4y \\ \frac{dt}{d\tau} &= 1 \end{aligned} \tag{2.6}$$

defined over $\mathcal{R}^3$ with respect to the coordinates (x, y, t). The integral curve of this extended vector field which passes through the point $(x_0, y_0, 0)$ when $\tau = 0$ is

$$\begin{aligned} s_x(\tau) &= e^{5\tau} x_0 \\ s_y(\tau) &= 3(e^{5\tau} - e^{4\tau})x_0 + e^{4\tau} y_0 \\ s_t(\tau) &= \tau. \end{aligned} \tag{2.7}$$

Comparing equations 2.5 and 2.7 it is clear that the image of the integral curve of the extended vector field 2.6 corresponds to the graph of equation 2.4.

Alternately, we could have chosen the extended vector field

$$\begin{aligned} \frac{dx}{d\tau} &= -5xt \\ \frac{dy}{d\tau} &= -3xt - 4yt \\ \frac{dt}{d\tau} &= -t. \end{aligned} \tag{2.8}$$

The integral curve of this vector field which passes through the point $(x_0, y_0, 1)$ when $\tau = 0$ is

$$\begin{aligned} s_x(\tau) &= e^{5e^{-\tau}} x_0 \\ s_y(\tau) &= 3(e^{5e^{-\tau}} - e^{4e^{-\tau}})x_0 + e^{4e^{-\tau}} y_0 \\ s_t(\tau) &= e^{-\tau}. \end{aligned} \tag{2.9}$$

Again, it is not too hard to see that the image of this integral curve locally corresponds with the graph of equation 2.4.

The only difference between this integral curve and the integral curve of equation 2.6 is the rate at which they flow along the graph of equation 2.4. In the first case, the time component moves at a constant rate; while in the second case, the time component goes to zero exponentially. $\diamond$

The point of this example is simply that vector fields are not the "right" geometric objects to model differential equations on the state-time space $\mathcal{R}^n \times \mathcal{R}$. On this extended space, all we really care about is the graph of the integral curve s_f, and this is not an integral curve; rather, it is a one-dimensional submanifold of $\mathcal{R}^n \times \mathcal{R}$. At each point $(s(\epsilon), \epsilon)$ the tangent space to this one-dimensional submanifold is a one-dimensional subspace, or a line, in $T_{(s_f(t),t)}(\mathcal{R}^n \times \mathcal{R})$. Consequently, each vector field on $\mathcal{R}^n$ maps to a unique distribution on $\mathcal{R}^n \times \mathcal{R}$.

As an alternative to modeling the differential equation using the vector field equation 2.1, we can instead model the differential equation in its Pfaffian form

$$\begin{aligned} dx^1 - f^1(x)dt &= 0 \\ &\vdots \\ dx^n - f^n(x)dt &= 0. \end{aligned} \tag{2.10}$$

These equations are defined pointwise on the state-time space, and should be thought of as representing a set of constraint conditions on the components of the tangent vectors at the corresponding point. If we represent an arbitrary tangent vector at the point $(x,t) \in \mathcal{R}^n \times \mathcal{R}$ by

$$v_{(x,t)} = \sum_{i=1}^{n} a^i \frac{\partial}{\partial x^i} + a^t \frac{\partial}{\partial t},$$

then the constraint conditions 2.10 require that

$$\begin{aligned} a^1 - f^1(x)a^t &= 0 \\ &\vdots \\ a^n - f^n(x)a^t &= 0. \end{aligned} \tag{2.11}$$

These constraint conditions are equivalent to the condition that

$$v_{(x,t)} = a^t \left(\sum_{i=1}^{n} f^i(x) \frac{\partial}{\partial x^i} + 1 \frac{\partial}{\partial t} \right)$$

which is exactly the requirement that $v_{(x,t)}$ be a pointwise scalar multiple of the extended vector field 2.2. Therefore, any vector field which satisfies the equations 2.10 will also produce a flow whose image coincides with the graph of s_f. Consequently, the solutions to the Pfaffian system form a distribution

$$\text{span} \left\{ \sum_{i=1}^{n} f^i(x) \frac{\partial}{\partial x^i} + 1 \frac{\partial}{\partial t} \right\}$$

whose integral submanifold passing through $(x_0, 0)$ coincides with the graph of s_f. Thus, the passage from vector fields on $\mathcal{R}^n$ to distributions on $\mathcal{R}^n \times \mathcal{R}$ is naturally induced by the equations 2.11.

2.2 Modeling Affine Nonlinear Control Systems on the State-Time Space

We next turn our attention to nonlinear control systems. Specifically, we want to consider affine nonlinear control systems defined over $\mathcal{R}^n$ by equations of the form

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u^i \tag{2.12}$$

where $f(x)$ is a smooth vector field called the drift, and the $g_i(x)$ are smooth vector fields which locally span an m-dimensional distribution G.

Equation 2.12 can be viewed in two different ways. First, we can look at this equation as parameterizing a family of vector fields. From this viewpoint, each feedback control law $u^i = \alpha^i(x,t)$ is equivalent to a time-varying vector field

$$\dot{x} = \hat{f}(x,t) = f(x) + \sum_{i=1}^{m} g_i(x) \alpha^i(x,t).$$

From the discussion in the previous section, we know that this vector field corresponds to the one-dimensional distribution

$$l_{(x,t)} = \text{span} \left\{ \sum_{j=1}^{m} \left(f^j(x) + \sum_{i=1}^{m} g_i^j(x) \alpha^i(x,t) \right) \frac{\partial}{\partial x^j} + 1 \frac{\partial}{\partial t} \right\}$$

defined over $\mathcal{R}^n \times \mathcal{R}$. Therefore, each feedback control law is equivalent to a particular one-dimensional distribution on the state-time space.

Second, we can look at equation 2.12 as a pointwise parameterization of the subset of all tangent vectors v_x at a point $x \in \mathcal{R}^n$ which satisfy the equation

$$v_x = f(x) + \sum_{i=1}^{m} g_i(x)c^i \tag{2.13}$$

for some arbitrary set of constants $c^1, \ldots, c^n$. Viewed in this way, equation 2.12 is seen to parameterize an affine plane of the same dimension as G in each tangent space $T_x\mathcal{R}^n$. Each vector v_x which satisfies equation 2.13 corresponds to the one-dimensional subspace

$$l_{(x,t)} = \text{span}\left\{ \sum_{j=1}^{m} \left(f^j(x) + \sum_{i=1}^{m} g_i^j(x)c^i \right) \frac{\partial}{\partial x^j} + 1\frac{\partial}{\partial t} \right\}$$

on $\mathcal{R}^n \times \mathcal{R}$. As the control parameters change, this subspace will change. The value of u therefore parameterizes a family of one-dimensional subspaces in the tangent space at (x,t). Moreover, each of these one-dimensional subspaces will lie in the $m+1$ dimensional subspace

$$F = \text{span}\left\{ \sum_{j=1}^{m} g_1^j(x)\frac{\partial}{\partial x^j}, \ldots, \sum_{j=1}^{m} g_m^j(x)\frac{\partial}{\partial x^j}, \left(\sum_{j=1}^{m} f^j(x)\frac{\partial}{\partial x^j} + 1\frac{\partial}{\partial t} \right) \right\}. \tag{2.14}$$

Thus we can think of the affine control system on $\mathcal{R}^n \times \mathcal{R}$ restricted to the point p as consisting of the collection of one-dimensional subspaces contained in the subspace F_p. Note that it is important to think of the choice of control input as specifying a whole subspace, rather than a particular vector field. Figure 2.2 illustrates these ideas. The picture on the left shows the tangent vectors $f(p)$ and $g(p)$ in the tangent space of a fixed state $p \in \mathcal{R}^2$. The dotted line shows the locus of points in this tangent space which can be generated using the control input. The picture on the right represents the corresponding tangent space above the fixed point $(p,t) \in \mathcal{R}^2 \times \mathcal{R}$. Each of the lines shown in this tangent space can be generated using the control input. Note that all of these lines are contained in the two dimensional subspace defined by

$$\text{span}\left\{ \begin{bmatrix} g(p) \\ 0 \end{bmatrix}, \begin{bmatrix} f(p) \\ 1 \end{bmatrix} \right\}.$$

The collection of all one-dimensional subspaces of $T_{(x,t)}(\mathcal{R}^n \times \mathcal{R})$ is called a projectivization of $T_{(x,t)}(\mathcal{R}^n \times \mathcal{R})$, and its properties are well known. In fact, we can construct a space consisting of all subspaces of $T_{(x,t)}(\mathcal{R}^n \times \mathcal{R})$ of any dimension $k \leq n$. Such an object is called a Grassmann Manifold and its properties will be discussed in Chapter 5.

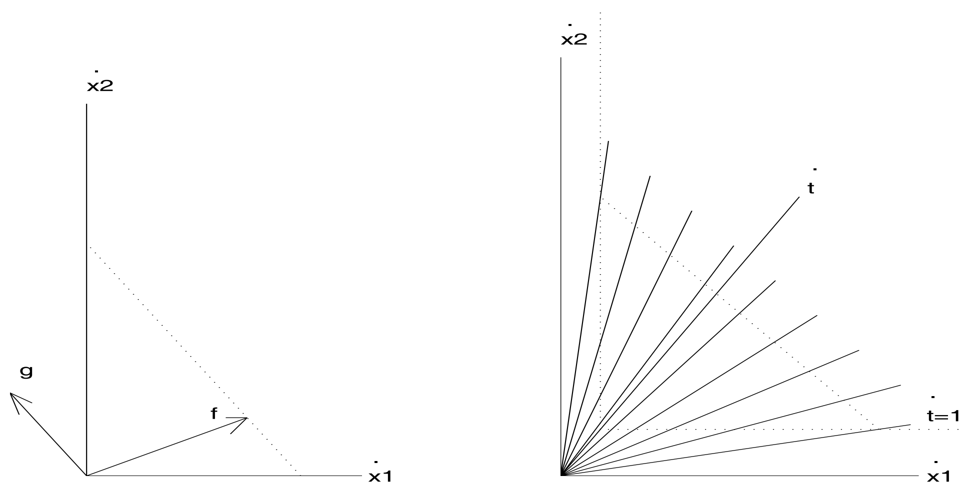

Figure 2.2: The Fibres of an Affine Control System

Part II

The Mathematics of Exterior Differential Systems

Chapter 3

Introduction

The purpose of this part of the dissertation is to develop the mathematical structures which will be used later in the dissertation. The reader is assumed to have an understanding of basic differential geometric structures: manifolds, coordinate charts, tangent spaces, etc., and basic algebraic structures: groups, rings, vector spaces, algebras, ideals, etc. A review of these topics is included in the appendices.

The material in this part of the dissertation is essentially presented twice. The first time through, all the constructions are developed relative to the familiar space $\mathcal{R}^n$. Next, we introduce the notion of a fibre bundle, and show how the constructions which were defined relative to $\mathcal{R}^n$ induce similar constructions over the tangent bundle of a smooth manifold.

Much of the material in this part of the dissertation originated in a course taught at U.C. Berkeley by Professor Sastry in the fall of 1994. As a class project, George Pappas, Chris Gerdes, and I prepared an expanded copy of the course notes on Exterior Differential Systems which was published as an Electronics Research Laboratory memo [17]. In turn, much of the material in this memo was based on the books by Munkres [5], Spivak [6], and Bryant, et al. [8]. Other useful sources are the book by Yang [7] and the book by Flanders [10]. In this reincarnation of the material, I have tried to put much more emphasis on the geometric aspects of the subject and their relation to the algebraic structures. In particular, I have added a chapter on Grassmann manifolds and a chapter on fibre bundles.

Chapter 4

Tensors and Forms

This chapter introduces tensors and forms. These objects are basic building blocks for the bundle structures which will be developed in the later chapters of Part II. The chapter is divided into five sections. The first section discusses the notion of duality and defines the dual space associated with a vector space. The second section presents the definition of a tensor as a real-valued multilinear function. This section also introduces a bilinear operator called the tensor product. The third section discusses alternating tensors and introduces the wedge product and interior product. The fourth section discusses the transformation which is induced on a tensor space by a linear transformation on the vector space over which the tensor space is defined. Finally, the fifth section concludes the chapter with a discussion of contravariant tensors.

4.1 The Dual Space of a Vector Space

Many of the ideas underlying the theory of multilinear algebra involve duality and the definition of the dual space of a vector space. Therefore, we will begin by briefly reviewing these concepts. We will only consider finite-dimensional vector spaces defined over $\mathcal{R}$. The dual space can also be defined for an infinite-dimensional vector space, but the theory becomes more involved.

Definition 4.1.1 *Let $(V, \mathcal{R})$ denote a vector space over $\mathcal{R}$. The dual space associated with $(V, \mathcal{R})$ is defined as the set of all linear mappings $f : V \rightarrow \mathcal{R}$. The dual space of V is denoted as V^* and the elements of V^* are called covectors.*

The next lemma gives the set V^* the structure of a vector space.

Lemma 4.1.1 *If for all $\alpha, \beta \in V^*$ and $c \in \mathcal{R}$, we define*

$$\begin{aligned} (\alpha+\beta)(v) &= \alpha(v)+\beta(v) \\ (c\alpha)(v) &= c \cdot \alpha(v), \end{aligned}$$

then V^* is a vector space over $\mathcal{R}$ with $dim(V^*) = dim(V)$. Furthermore, if we pick a set of basis vectors $\{v_1, \ldots, v_n\}$ for V, then the set of linear functions $\phi^i : V \rightarrow \mathcal{R}$, $1 \leq i \leq n$, defined by

$$\phi^i(v_j) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

form a basis of V^* called the dual basis.

Proof: See Munkres [5] page 220. □

Example: Let $V = \mathcal{R}^n$ with the standard basis $e_1, \ldots, e_n$, and let $\phi^1, \ldots, \phi^n$ be the dual basis. If

$$x \in \mathcal{R}^n = \sum_{j=1}^{n} x^j e_j,$$

then evaluating each function in the dual basis at x gives

$$\phi^i(x) = \phi^i(\sum_{j=1}^{n} x^j e_j) = \sum_{j=1}^{n} x^j \phi^i(e_j) = x^i.$$

Since the functions $\phi^1, \ldots, \phi^n$ form a basis for V^*, a general covector in $(\mathcal{R}^n)^*$ is of the form

$$f = a_1\phi^1 + \ldots + a_n\phi^n.$$

Evaluating this covector at x gives

$$f(x) = a_1 x^1 + \ldots + a_n x^n.$$

If we think of a vector as a column matrix and a covector as a row matrix, then

$$f(x) = [a_1 \ldots a_n] \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}$$

. ◇

For each subspace $W \subset V$, there is a corresponding "perpendicular" subspace $W^\perp \subset V^*$.

Definition 4.1.2 *Given a subspace $W \subset V$ its annihilator is the subspace $W^\perp \subset V^*$ defined by*

$$W^\perp := \{\alpha \in V^* \mid \alpha(v) = 0 \ \forall \ v \in W\}.$$

Given a subspace $X \subset V^*$, its annihilator is the subspace $X^\perp \subset V$ defined by

$$X^\perp := \{v \in V \mid \alpha(v) = 0 \ \forall \ \alpha \in X\}.$$

Given a linear mapping between any two vector spaces $F : V_1 \longrightarrow V_2$ we can define an induced linear mapping between their dual spaces.

Definition 4.1.3 *Given a linear mapping $F : V_1 \longrightarrow V_2$, its dual map is the linear mapping $F^* : V_2^* \longrightarrow V_1^*$ defined by*

$$(F^*(\alpha))(v) = \alpha(F(v)), \ \forall \ \alpha \in V_2^*, \ v \in V_1.$$

Since V^* is a vector space, it also has a dual space which is denoted as V^{**}. There exists a "natural" identification $i : V \longrightarrow V^{**}$ which is defined for all $v \in V$ and $\alpha \in V^*$ by

$$(i(v))(\alpha) = \alpha(v).$$

For all finite-dimensional vector spaces, this fact allows us to treat V and V^{**} as essentially the same object. For example, we could have defined the annihilator as

$$W^\perp := \{\alpha \in V^* \mid \alpha(v) = 0 \ \forall \ v \in W \subset V\}.$$

Using this definition, the annihilator of a subspace $W^* \subset V^*$ is defined as a subspace of V^{**}. However, we can use the natural identification to map this subspace back to V in which case we recover our original definition.

4.2 Multilinear Functions and Tensors

Let $V_1, \ldots, V_k$ be a collection of real vector spaces. A function

$$f : V_1 \times \ldots \times V_k \to \mathcal{R}$$

is said to be linear in the ith variable if the function $T : V_i \to \mathcal{R}$ defined with fixed $v_j \neq v_i$ as

$$T(v) = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_k)$$

is linear. A function is called multilinear if it is linear in each variable.

A multilinear function $T : V^k \to \mathcal{R}$ is said to be a covariant tensor of order k or simply a k-tensor. The set of all k-tensors on V is denoted $\mathcal{L}^k(V)$. For $k = 1$, we have $\mathcal{L}^k(V) = V^*$, the dual space of V. Therefore, we can think of covariant tensors as generalized covectors.

Examples:

1. A typical example of a multilinear function is the inner product of two vectors. From the definition of the inner product, we know that for any vectors $x, y, z \in \mathcal{R}^n$

$$< a \cdot x, y >=< x, a \cdot y >= a \cdot < x, y >$$

$$< x + z, y >=< x, y > + < x, z >=< x, y + z > .$$

2. Another important example of a multilinear function is the determinant. If $v_1, v_2, \ldots, v_n$ are n column vectors in $\mathcal{R}^n$, then

$$det[v_1 \ v_2 \ \ldots \ v_n]$$

is multilinear. This fact can be verified using a row or column expansion which expresses the determinant in terms of its minors.

$\diamond$

As in the case of V^*, each $\mathcal{L}^k(V)$ can be made into a vector space.

Lemma 4.2.1 *If for $S, T \in \mathcal{L}^k(V)$ and $c \in \mathcal{R}$ we define*

$$\begin{aligned} (S+T)(v_1, \ldots, v_k) &= S(v_1, \ldots, v_k) + T(v_1, \ldots, v_k) \\ (cT)(v_1, \ldots, v_k) &= c \cdot T(v_1, \ldots, v_k), \end{aligned}$$

then the set of all k-tensors on V, $\mathcal{L}^k(V)$, is a real vector space.

Proof: See Munkres [5] page 220. $\square$

Because of their multilinear structure, two tensors are equal if they agree on any set of basis elements.

Lemma 4.2.2 *Let $a_1, \ldots, a_n$ be a basis for V. Let $f, g : V^k \to \mathcal{R}$ be k-tensors on V. If*

$$f(a_{i_1}, \ldots, a_{i_k}) = g(a_{i_1}, \ldots, a_{i_k})$$

for every k-tuple (multi-index)

$$I = (i_1, \ldots, i_k) \in \{1, 2, \ldots, n\}^k,$$

then $f = g$.

Proof: See Munkres [5] page 221. □

Lemma 4.2.2 allows us to construct a basis for the space $\mathcal{L}^k(V)$.

Lemma 4.2.3 *Let $a_1, \ldots, a_n$ be a basis for V. Let $I = (i_1, \ldots, i_k) \in \{1, 2, \ldots, n\}^k$. Then there is a unique tensor ϕ^I on V such that for every k-tuple*

$$J = (j_1, \ldots, j_k) \in \{1, 2, \ldots, n\}^k,$$

$$\phi^I(a_{j_1}, \ldots, a_{j_k}) = \begin{cases} 0 & \text{if } I \neq J \\ 1 & \text{if } I = J, \end{cases}$$

and the collection of all the ϕ^I forms a basis for $\mathcal{L}^k(V)$.

Proof: Uniqueness follows from Lemma 4.2.2. To construct the functions ϕ^I, we start with a basis for V^*, $\phi^i : V \to \mathcal{R}$, defined by

$$\phi^i(a_j) = \delta_{ij}.$$

We then define each ϕ^I as

$$\phi^I = \phi^{i_1}(v_1) \cdot \phi^{i_2}(v_2) \cdot \ldots \cdot \phi^{i_k}(v_k)$$

and claim that these ϕ^I form a basis for $\mathcal{L}^k(V)$.

To show this, we select an arbitrary k-tensor $f \in \mathcal{L}^k(V)$ and define the scalars

$$\alpha_I := f(a_{i_1}, \ldots, a_{i_k}).$$

Next, we define a k-tensor

$$g = \sum_J \alpha_J \phi^J$$

where

$$J \in \{1, \ldots, n\}^k.$$

Then by Lemma 4.2.2, $f \equiv g$. □

Since there are n^k distinct k-tuples from the set $\{1, \ldots, k\}$, the space $\mathcal{L}^k(V)$ has dimension n^k.

Example: Let $V = \mathcal{R}^n$ with the standard basis $e_1, \ldots, e_n$, and let $\phi^1, \ldots, \phi^n$ be the corresponding dual basis. If

$$x \in \mathcal{R}^n = \sum_{j=1}^{n} x^j e_j,$$

then evaluating each function in the dual basis at x gives

$$\phi^i(x) = \phi^i(\sum_{j=1}^{n} x^j e_j) = \sum_{j=1}^{n} x^j \phi^i(e_j) = x^i.$$

Likewise, we can let $I = (i_1, \ldots, i_k)$ and $v_k = \sum_{i=1}^{n} x_k^i e_i$. Evaluating the basis vectors for $\mathcal{L}^k(V)$ at $(v_1, \ldots, v_k)$ gives

$$\begin{aligned} \phi^I(v_1, \ldots, v_k) &= \phi^{i_1}(v_1) \cdot \phi^{i_2}(v_2) \cdot \ldots \cdot \phi^{i_k}(v_k) \\ &= x_1^{i_1} \cdot x_2^{i_2} \cdot \ldots \cdot x_k^{i_k}. \end{aligned}$$

Since the tensors $\phi^1, \ldots, \phi^n$ form a basis for V^*, evaluating a general 1-tensor $f \in (\mathcal{R}^n)^*$ at $\sum_{i=1}^{n} x^i e_i \in \mathcal{R}^n$ gives

$$f(x) = a_1 x^1 + \ldots + a_n x^n.$$

Evaluating a general 2-tensor at $(x, y) \in \mathcal{R}^2$ gives

$$g(v_1, v_2) = \sum_{i,j=1}^{n} a_{ij} x_1^i x_2^j$$

and evaluating a general k-tensor at $(v_1, \ldots, v_k) \in \mathcal{R}^k$ gives

$$g(v_1, v_2, \ldots, v_k) = \sum_{i_1,\ldots,i_k=1}^{n} a_{i_1,\ldots,i_k} x_1^{i_1} \ldots x_k^{i_k}.$$

$\diamond$

4.2.1 Tensor Products

We now introduce a product operation into the set of all tensors on V and outline its basic properties.

Definition 4.2.1 *Let $f \in \mathcal{L}^k(V)$ and $g \in \mathcal{L}^l(V)$. The tensor product $f \otimes g$ of f and g is a tensor in $\mathcal{L}^{k+l}(V)$ and is defined by*

$$(f \otimes g)(v_1, \ldots, v_{k+l}) := f(v_1, \ldots, v_k) \cdot g(v_{k+1}, \ldots, v_{k+l}).$$

Lemma 4.2.4 *Let f, g, h be tensors on V and $c \in \mathcal{R}$. Then we have*

1. *Associativity* $f \otimes (g \otimes h) = (f \otimes g) \otimes h$
2. *Homogeneity* $cf \otimes g = c(f \otimes g) = f \otimes cg$
3. *Distributivity* $(f + g) \otimes h = f \otimes h + g \otimes h$
4. *Given a basis $a_1, \ldots, a_n$ for V, the basis tensors satisfy* $\phi^I = \phi^{i_1} \otimes \phi^{i_2} \otimes \cdots \otimes \phi^{i_k}$

Proof: See Munkres [5] page 224. $\square$

We can also define the tensor product of two subspaces $U, W \subset V^*$ as

$$U \otimes W := span\{x \in \mathcal{L}^2(V) \mid x = u \otimes w,\ u \in U,\ w \in W\}.$$

Therefore, from Lemma 4.2.3 we can conclude that

$$V^* \otimes V^* = \mathcal{L}^2(V).$$

More generally we have

$$\underbrace{V^* \otimes \ldots \otimes V^*}_{k-times} = \bigotimes^{k} V^* = \mathcal{L}^k(V)$$

4.3 Alternating Multilinear Functions and Forms

In this section we introduce the concept of an alternating tensor. In order to do this, we need to know some facts about permutations.

4.3.1 Permutations

Definition 4.3.1 *A permutation of the set of integers $\{1, 2, \ldots, k\}$ is a one-to-one function σ mapping this set onto itself.*

The set of all permutations σ is a group under function composition called the symmetric group on $\{1, \ldots, k\}$ and is denoted by S_k. Permutations simply reshuffle the elements of a finite set. As a result, the number of permutations in S_k is $k!$.

Definition 4.3.2 *Given $1 \leq i < k$, a permutation e_i is called elementary if given some $i \in \{1, 2, \ldots, k\}$ we have*

$$\begin{aligned} e_i(j) &= j \qquad for \qquad j \neq i, i+1 \\ e_i(i) &= i+1 \\ e_i(i+1) &= i \end{aligned}$$

An elementary permutation leaves the set intact except for consecutive elements i and $i+1$ which are switched. The space S_k can be constructed from the elementary permutations.

Lemma 4.3.1 *Every permutation $\sigma \in S_k$ can be written as the composition of elementary permutations.*

Proof: See Munkres [5] page 227. □

Definition 4.3.3 *Let $\sigma \in S_k$. Consider the set of all pairs of integers i, j from the set $\{1, \ldots, k\}$ for which $i < j$ and $\sigma(i) > \sigma(j)$. Each such pair is called an inversion in σ. The sign of σ is defined to be the number -1 if the number of inversions is odd and $+1$ if it is even. We call σ an odd or even permutation respectively. The sign of σ is denoted by $sgn(\sigma)$.*

The following lemma helps us calculate the sign of permutations.

Lemma 4.3.2 *Let $\sigma, \tau \in S_k$. Then*

1. *If σ equals the composite of m elementary permutations, then $sgn(\sigma) = (-1)^m$*
2. $sgn(\sigma \circ \tau) = sgn(\sigma) \cdot sgn(\tau)$
3. $sgn(\sigma^{-1}) = sgn(\sigma)$
4. *If $p \neq q$, and if τ is the permutation that exchanges p and q and leaves all other integers fixed, then $sgn(\tau) = -1$*

Proof: See Munkres [5] page 228. □

4.3.2 Forms

We are now ready to define forms.

Definition 4.3.4 *Let f be an arbitrary k-tensor on V. If σ is a permutation of $\{1, \ldots, k\}$, we define f^σ by the equation*

$$f^\sigma(v_1, \ldots, v_k) = f(v_{\sigma(1)}, \ldots, v_{\sigma(k)}). \tag{4.1}$$

Since f is linear in each of its variables, so is f^σ. The tensor f is said to be symmetric if $f = f^e$ for each elementary permutation e, and it is said to be alternating if $f = -f^e$ for every elementary permutation e.

In other words, f is symmetric if for all i

$$f(v_1, \ldots, v_i, v_{i+1}, \ldots, v_k) = f(v_1, \ldots, v_{i+1}, v_i, \ldots, v_k) \tag{4.2}$$

and alternating if

$$f(v_1, \ldots, v_i, v_{i+1}, \ldots, v_k) = -f(v_1, \ldots, v_{i+1}, v_i, \ldots, v_k). \tag{4.3}$$

A real-valued alternating k-tensor is called a $\underline{k\text{-form}}$. We will denote the set of all k-forms on V by $\Lambda^k(V^*)$. The reason for this notation will be apparent when we introduce the wedge product in the next section.

One can verify that the set of all k-forms is closed under addition and scalar multiplication. Therefore, $\Lambda^k(V^*)$ is a linear subspace of the space $\mathcal{L}^k(V)$ of all k-tensors on V.

In the special case of $\mathcal{L}^1(V)$, elementary permutations cannot be performed and therefore every 1-tensor is vacuously alternating. Therefore $\Lambda^1(V^*) = \mathcal{L}^1(V) = V^*$. Furthermore, for completeness, we define $\Lambda^0(V^*) = \mathcal{R}$.

Examples: Elementary tensors are not alternating but the following linear combination

$$f = \phi^i \otimes \phi^j - \phi^j \otimes \phi^i$$

is alternating. To see this, let $V = \mathcal{R}^n$ and let ϕ^i be the usual dual basis. Then

$$f(x, y) = x_i y_j - x_j y_i = det \begin{bmatrix} x_i & y_i \\ x_j & y_j \end{bmatrix}$$

and it is easily seen that $f(x, y) = -f(y, x)$. Similarly, the function

$$g(x, y, z) = det \begin{bmatrix} x_i & y_i & z_i \\ x_j & y_j & z_j \\ x_k & y_k & z_k \end{bmatrix}$$

is an alternating 3-tensor. $\diamond$

We are interested in obtaining a basis for the linear space $\Lambda^k(V^*)$. We start with the following lemma.

Lemma 4.3.3 *Let f be a k-tensor on V and $\sigma, \tau \in S_k$. Then*

1. *The transformation $f \longrightarrow f^\sigma$ is a linear transformation from $\mathcal{L}^k(V^*)$ to $\mathcal{L}^k(V^*)$. It has the property that for all $\sigma, \tau \in S_k$,*
$$(f^\sigma)^\tau = f^{\tau \circ \sigma}.$$
2. *The tensor f is alternating if and only if $f^\sigma = sgn(\sigma) \cdot f$ for all $\sigma \in S_k$.*
3. *If f is alternating and if $v_p = v_q$ with $p \neq q$, then $f(v_1, \ldots, v_k) = 0$.*

Proof: The linearity property is obvious since $(af + bg)^\sigma = af^\sigma + bg^\sigma$. Furthermore,

$$\begin{aligned} (f^\sigma)^\tau(v_1, \ldots, v_k) &= f^\sigma(v_{\tau(1)}, \ldots, v_{\tau(k)}) \\ &= f^\sigma(w_1, \ldots, w_k) \qquad w_i = v_{\tau(i)} \\ &= f(w_{\sigma(1)}, \ldots, w_{\sigma(k)}) \\ &= f(v_{\tau(\sigma(1))}, \ldots, v_{\tau(\sigma(k))}) \\ &= f^{\tau\circ\sigma}(v_1, \ldots, v_k). \end{aligned}$$

Let σ be an arbitrary permutation. We can write it as

$$\sigma = \sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_m$$

where each σ_i is an elementary permutation, so

$$\begin{aligned} f^\sigma &= f^{\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_m} \\ &= ((\ldots(f^{\sigma_m})\ldots)^{\sigma_2})^{\sigma_1} \\ &= (-1)^m \cdot f \\ &= sgn(\sigma) \cdot f. \end{aligned}$$

Finally, suppose $v_p = v_q$ and $p \neq q$. Let τ be a permutation that exchanges p and q. Since $v_p = v_q$,

$$f^\tau(v_1, \ldots, v_k) = f(v_1, \ldots, v_k),$$

and since τ is an alternating tensor,

$$f^\tau(v_1, \ldots, v_k) = sgn(\tau) \cdot f(v_1, \ldots, v_k) = -f(v_1, \ldots, v_k).$$

Therefore, $f(v_1, \ldots, v_k) = 0$. □

As a result of Lemma 4.3.3, if $k > n$, then the space $\Lambda^k(V^*)$ is trivial since one of the basis elements must appear in the k-tuple more than once. Hence, for $k > n$, $\Lambda^k(V^*) = 0$. We have also seen that for $k = 1$ we have $\Lambda^1(V^*) = \mathcal{L}^1(V) = V^*$, so one can use the dual basis as a basis for $\Lambda^1(V^*)$. We are therefore left with the cases where $1 < k \leq n$. The key argument here is that in order to specify a form we simply need to define it on an ascending k-tuple of basis elements since, by applying Lemma 4.3.3, every other combination can be obtained by permuting the k-tuple.

Lemma 4.3.4 *Let $a_1, a_2, \ldots, a_n$ be a basis for V. If f, g are forms on V and if*

$$f(a_1, a_2, \ldots, a_n) = g(a_1, a_2, \ldots, a_n)$$

for every ascending k-tuple of integers $\{1, 2, \ldots, k\}$, then $f = g$.

Proof: See Munkres [5] page 231. □

Lemma 4.3.5 *Let $a_1, \ldots, a_n$ be a basis for V. Let $I = (i_1, \ldots, i_k) \in \{1, 2, \ldots, n\}^k$ be an ascending k-tuple. There is a unique form ψ^I on V such that for every ascending k-tuple $J = (j_1, \ldots, j_k) \in \{1, 2, \ldots, n\}^k$,*

$$\psi^I(a_{j_1}, \ldots, a_{j_k}) = \begin{cases} 0 & \text{if } I \neq J \\ 1 & \text{if } I = J. \end{cases}$$

The forms ψ^I form a basis for $\Lambda^k(V^*)$. The forms ψ^I also satisfy the formula

$$\psi^I = \sum_{\sigma \in S_k} sgn(\sigma)(\phi^I)^\sigma.$$

Proof: See Munkres [5] pages 232-233. □

The forms ψ^I are called the elementary forms on V corresponding to the basis $a_1, \ldots, a_n$ for V. Therefore, every k-form f may be uniquely expressed by

$$f = \sum_J d_J \psi^J$$

where J indicates that summation extends over all ascending k-tuples.

Since $\Lambda^1(V^*)$ is isomorphic to V^*, it must have dimension n. In order to determine the dimension of $\Lambda^k(V^*)$ when $k > 1$, we need to find the number of possible ascending k-tuples from the set $\{1, 2, \ldots, n\}$. If we choose k elements from a set of n elements, there is only one way to put them in ascending order. Therefore the number of ascending k-tuples and the dimension of $\Lambda^k(V^*)$ is

$$dim(\Lambda^k(V^*)) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

4.3.3 The Wedge Product

Just as we defined the tensor product operation in the set of all tensors on a vector space V, we can define an analogous product operation, the wedge product, in the space of all forms. The tensor product alone does not suffice since even if $f \in \Lambda^k(V^*)$ and $g \in \Lambda^l(V^*)$ are alternating, their tensor product $f \otimes g \in \mathcal{L}^{k+l}(V)$ need not be alternating. We therefore construct an alternating operator taking k-tensors to k-forms.

Lemma 4.3.6 *For any tensor $f \in \mathcal{L}^k(V)$, define $Alt : \mathcal{L}^k(V) \to \Lambda^k(V^*)$ by*

$$Alt(f) = \frac{1}{k!} \sum_{\sigma \in S_k} sgn(\sigma) f^\sigma.$$

Then $Alt(f) \in \Lambda^k(V^*)$ and if $f \in \Lambda^k(V^*)$, then $Alt(f) = f$.

Proof: The fact that $Alt(f) \in \Lambda^k(V^*)$ is a consequence of Lemma 4.3.3 parts (1) and (2). Simply expanding the summation for $f \in \Lambda^k(V^*)$ yields that $Alt(f) = f$. □

Example: Let $f(x, y)$ be any 2-tensor. Applying the alternating operator, we obtain

$$Alt(f) = \frac{1}{2}(f(x, y) - f(y, x))$$

which is clearly alternating. Similarly, for any 3-tensor g(x,y,z), we have

$$Alt(g) = \frac{1}{6}(g(x, y, z) + g(y, z, x) + g(z, x, y) - g(y, x, z) - g(z, y, x) - g(x, z, y)).$$

◇

Definition 4.3.5 *Given $f \in \Lambda^k(V^*)$ and $g \in \Lambda^l(V^*)$, we define the wedge or exterior product, $f \wedge g \in \Lambda^{k+l}(V^*)$, by the equation*

$$f \wedge g = \frac{(k+l)!}{k!l!} Alt(f \otimes g).$$

Therefore, given two forms, the wedge product first obtains the tensor product of the two forms, then uses the alternating operator in order to obtain a new form, and finally normalizes it. There are two reasons for the somewhat complicated normalization constant. The first reason is so that if f is alternating then $Alt(f) = f$. The second reason is that we want the wedge product to be associative. The normalizing coefficient ensures both properties. Since forms of order zero are elements of $\mathcal{R}$, we define the wedge product of an alternating 0-tensor and any alternating k-tensor to be the usual multiplication. The following lemma lists some important properties of the wedge product.

Lemma 4.3.7 *Let $f \in \Lambda^k(V^*)$, $g \in \Lambda^l(V^*)$, and $h \in \Lambda^m(V^*)$. Then*

1. *Associativity* $f \wedge (g \wedge h) = (f \wedge g) \wedge h$

2. *Homogeneity* $cf \wedge g = c(f \wedge g) = f \wedge cg$

3. *Distributivity* $(f+g) \wedge h = f \wedge h + g \wedge h$
$h \wedge (f+g) = h \wedge f + h \wedge g$

4. *Skew-commutativity* $g \wedge f = (-1)^{kl} f \wedge g$

Proof: Properties (2), (3), and (4) follow directly from the definitions of the alternating operator and the tensor product. Associativity, property (1), requires a few more manipulations (see Spivak [13] pages 80-81). □

Example: Let $f(x) \in \Lambda^1(V^*)$ and $g(y,z) \in \Lambda^2(V^*)$. Then

$$\begin{aligned} f \wedge g &= \frac{(2+1)!}{2!1!}\frac{1}{3!}(f(x) \otimes g(y,z) + f(y) \otimes g(z,x) + \\ &+ f(z) \otimes g(x,y) - f(y) \otimes g(x,z) - f(z) \otimes g(y,x) - f(x) \otimes g(z,y)) \end{aligned}$$

We can also check that

$$f \wedge f = \frac{(1+1)!}{1!1!}\frac{1}{2!}(f(x) \otimes f(x) - f(x) \otimes f(x)) = 0$$

which can also been seen from the skew-commutativity of exterior multiplication. ◇

We can now formulate a basis for $\Lambda^k(V^*)$ more elegantly in terms of the dual basis for V.

Lemma 4.3.8 *Given a basis $a_1, \ldots, a_n$ for a vector space V, let $\phi^1, \ldots, \phi^n$ denote its dual basis, and let ψ^I denote an element in the corresponding set of elementary k-forms. If $I = (i_1, \ldots, i_k)$ is any ascending k-tuple of integers, then*

$$\psi^I = \phi^{i_1} \wedge \phi^{i_2} \wedge \ldots \wedge \phi^{i_k}.$$

Proof: May be deduced from the construction of the elementary k-forms in Lemma 4.3.5. □

By Lemma 4.3.8, any k-form $f \in \Lambda^k(V^*)$ may be expressed in terms of the dual basis $\phi^1, \ldots, \phi^n$ as

$$f = \sum_J d_{j_1,\ldots,j_k} \phi^{j_1} \wedge \phi^{j_2} \wedge \ldots \wedge \phi^{j_k} \tag{4.4}$$

for all ascending k-tuples $J = (j_1, \ldots, j_k)$ and some scalars, $d_{j_1,\ldots,j_k}$. If we require the coefficients to be skew-symmetric, then

$$d_{i_1,\ldots,i_l,i_{l+1},\ldots,i_k} = -d_{i_1,\ldots,i_{l+1},i_l,\ldots,i_k},\ \forall\, l \in \{1,\ldots,k-1\},$$

and we can extend this summation over all k-tuples

$$f = \frac{1}{k!}\sum_{i_1,\ldots,i_k=1}^{n} d_{i_1,\ldots,i_k} \phi^{i_1} \wedge \phi^{i_2} \wedge \ldots \wedge \phi^{i_k}. \tag{4.5}$$

The wedge product has a number of nice properties which make it a useful algebraic tool. For example, the wedge product provides a way to check whether a set of 1-forms is linearly independent.

Lemma 4.3.9 *If $\omega^1, \ldots, \omega^k$ are 1-forms over V then*

$$\omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k = 0$$

if and only if $\omega^1, \ldots, \omega^k$ are linearly dependent.

Proof: Suppose that the 1-forms $\omega^1, \ldots, \omega^k$ are linearly independent, and pick $\alpha^{k+1}, \ldots, \alpha^n$ to complete a basis for V^*. From Lemma 4.3.8, we know that

$$\omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k$$

is a basis element for $\Lambda^k(V^*)$. Therefore, it must be nonzero.

If the 1-forms $\omega^1, \ldots, \omega^k$ are linearly dependent, then at least one of them can be written as a linear combination of the rest. Without loss of generality, assume that ω^k is linearly dependent. We then have

$$\omega^k = \sum_{i=1}^{k-1} c_i \omega^i.$$

Furthermore, the skew-commutativity of the wedge product implies that

$$\omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k = \omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^{k-1} \wedge (\sum_{i=1}^{k-1} c_i \omega^i) = 0$$

□

This result allows us to give a geometric interpretation to a nonzero k-form

$$\omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k \neq 0$$

by associating it with the subspace

$$W := span\{\omega^1, \ldots, \omega^k\} \subset V^*.$$

An obvious question which arises is what happens if we select a different basis for W.

Lemma 4.3.10 *Given a subspace $W \subset V^*$ and two sets of* 1*-tensors which span W,*

$$W = span\{\omega^1, \ldots, \omega^k\} = span\{\alpha^1, \ldots, \alpha^k\},$$

there exists a nonzero scalar $c \in \mathcal{R}$ such that

$$c \cdot \omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k = \alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k \neq 0$$

Proof: Each α^i can be written as a linear combination of the ω^i

$$\alpha^i = \sum_{j=1}^{k} a^i_j \omega^j.$$

Therefore, the product

$$\alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k = (\sum_{j=1}^{k} a^j_1 \omega^j) \wedge \ldots \wedge (\sum_{j=1}^{k} a^j_k \omega^j).$$

Multiplying this out gives

$$\alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k = \sum_{i_1, \ldots, i_k = 1}^{n} b_{i_1, \ldots, i_k} \omega^{i_1} \wedge \omega^{i_2} \wedge \ldots \wedge \omega^{i_k}.$$

Finally, using Lemma 4.3.9 and the skew-commutativity of the wedge product, we get

$$c \cdot \omega^1 \wedge \omega^2 \wedge \ldots \wedge \omega^k = \alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k \neq 0.$$

□

Therefore, the k-fold wedge product of all sets of linearly independent 1-forms which spans a subspace of $W \subset V^*$ differ by only a scalar constant. We can therefore define an equivalence class of basis sets for W.

Definition 4.3.6 *Let $\xi = x^1 \wedge \ldots \wedge x^k$. We define an equivalence class*

$$[x^1 \ldots x^k] := \{\pi \in \Lambda^k(V^*) \mid \pi = c \cdot \xi, \textit{ for some nonzero } c \in \mathcal{R}\}$$

called the Grassmann coordinate *of ξ.*

The set of all such equivalence classes can be put in one-to-one correspondence with the set of all k-dimensional subspaces of V^*. This set of subspaces is called the Grassmann manifold of k-planes in V^* and is denoted as $G_k^{V^*}$.

Definition 4.3.7 *A k-form $\xi \in \Lambda^k(V^*)$ is* decomposable *if there exist*

$$x^1, x^2, \ldots, x^k \in \Lambda^1(V^*)$$

such that $\xi = x^1 \wedge x^2 \wedge \ldots \wedge x^p$.

There exist k-forms which are not decomposable. To see this, consider the following example.

Example: Let $\xi = \phi^1 \wedge \phi^2 + \phi^3 \wedge \phi^4 \in \Lambda^2((\mathcal{R}^4)^*)$. If ξ is decomposable, then we must have $\xi \wedge \xi = 0$. The reason for this is that if ξ can be expressed as

$$\xi = \alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k,$$

then it follows that

$$\xi \wedge \xi = \alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k \wedge \alpha^1 \wedge \alpha^2 \wedge \ldots \wedge \alpha^k = 0.$$

In this case, we have

$$\xi \wedge \xi = 2\phi^1 \wedge \phi^2 \wedge \phi^3 \wedge \phi^4 \neq 0.$$

Therefore, ξ is not decomposable. Notice that this is a necessary but not sufficient condition and, in particular, it does not apply to odd-dimensional forms. ◇

Even if a k-form ξ is not decomposable, it may still be possible to factor out a 1-form from every term in the summation which defines it.

Example: Let $\xi = \phi^1 \wedge \phi^2 \wedge \phi^5 + \phi^3 \wedge \phi^4 \wedge \phi^5 \in \Lambda^3((\mathcal{R}^5)^*)$. From the previous example, we know that this form is not decomposable, but the 1-form ϕ^5 can clearly be factored from every term

$$\xi = (\phi^1 \wedge \phi^2 + \phi^3 \wedge \phi^4) \wedge \phi^5 = \hat{\xi} \wedge \phi^5$$

◇

Definition 4.3.8 *Let $\xi \in \Lambda^k(V^*)$. We define a subspace $L_\xi \subset V^*$*

$$L_\xi := \{\omega \in V^* \mid \xi = \hat{\xi} \wedge \omega \textit{ for some } \hat{\xi} \in \Lambda^{k-1}(V^*)\}$$

called the divisor space *of ξ. Any $\omega \in L_\xi$ is called a* divisor *of ξ.*

Lemma 4.3.11 *A 1-form $\omega \in V^*$ is a divisor of $\xi \in \Lambda^k(V^*)$ if and only if*

$$\omega \wedge \xi \equiv 0.$$

Proof: Pick a basis $\phi^1, \phi^2, \ldots, \phi^n$ for V^* such that $\omega = \phi^1$. With respect to this basis, ξ can be written as

$$\xi = \sum_J^n d_{j_1,\ldots,j_k} \phi^{j_1} \wedge \phi^{j_2} \wedge \ldots \wedge \phi^{j_k} \tag{4.6}$$

for all ascending k-tuples $J = (j_1, \ldots, j_k)$ and some scalars, $d_{j_1,\ldots,j_k}$. If ω is a divisor of ξ, then it must be contained in each nonzero term of this summation. Therefore $\omega \wedge \xi$ must be identically 0.

If $\omega \wedge \xi \equiv 0$, then every nonzero term of ξ must contain ω. Otherwise, we would have for $j_1, \ldots, j_k \neq 1$,

$$\omega \wedge \phi^{j_1} \wedge \ldots \wedge \phi^{j_k} = \phi^1 \wedge \phi^{j_1} \wedge \ldots \wedge \phi^{j_k}$$

which is a basis element of $\Lambda^{k+1}(V^*)$ and therefore nonzero. This contradicts the assumption $\omega \wedge \xi \equiv 0$. □

If we select a basis $\phi^1, \phi^2, \ldots, \phi^n$ for V^* such that

$$span\{\phi^1, \phi^2, \ldots, \phi^l\} = L_\xi,$$

then ξ can be written as

$$\xi = \hat{\xi} \wedge \phi^1 \wedge \ldots \wedge \phi^l$$

where $\hat{\xi} \in \Lambda^{k-l}(V^*)$ is not decomposable and involves only the one-forms $\phi^{l+1}, \ldots, \phi^n$.

4.3.4 The Interior Product

A second useful operation on tensors is called the interior product.

Definition 4.3.9 *The* <u>*interior product*</u> *is a linear mapping* $\lrcorner : V \times \mathcal{L}^k(V) \to \mathcal{L}^{k-1}(V)$ *which operates on a vector* $v \in V$ *and a tensor* $T \in \mathcal{L}^k(V)$ *and produces a tensor* $(v \lrcorner T) \in \mathcal{L}^{k-1}(V)$ *defined by*

$$(v \lrcorner T)(v_1, \ldots, v_{k-1}) := T(v, v_1, \ldots, v_{k-1}).$$

The interior product has the following properties.

Lemma 4.3.12 *Let* a, b, c, d *be real numbers and* $v, w \in V$, $g, h \in \mathcal{L}^l(V)$, $r \in \Lambda^s(V^*)$, *and* $f \in \Lambda^m(V^*)$. *Then we have*

1. *Bilinearity* $$\begin{aligned} (av + bw) \lrcorner g &= a(v \lrcorner g) + b(w \lrcorner g) \\ v \lrcorner (cg + dh) &= c(v \lrcorner g) + d(v \lrcorner h) \end{aligned}$$

2. $v \lrcorner (f \wedge r) = (v \lrcorner f) + (-1)^m f \wedge (v \lrcorner g)$

Proof: See Yang [7] page 12. □

The next result illustrates a useful property of the interior product.

Lemma 4.3.13 *Let* $a_1, \ldots, a_n$ *be a basis for* V. *Then the value of a* k*-form* $\omega \in \Lambda^k(V^*)$ *is independent of a basis element* a_i *if and only if* $a_i \lrcorner \omega \equiv 0$.

Proof: Let $\phi^1, \ldots, \phi^n$ be the dual basis to $a_1, \ldots, a_n$. Then ω can be written with respect to the dual basis as

$$\omega = \sum_J d_J \phi^{j_1} \wedge \phi^{j_2} \wedge \ldots \wedge \phi^{j_k} = \sum_J d_J \psi^J$$

where the sum is taken oven all ascending k-tuples J. If a basis element ψ^J does not contain ϕ^i, then clearly

$$a_i \lrcorner \psi^J \equiv 0.$$

If a basis element contains ϕ^i, then

$$a_i \lrcorner \ \phi^{j_1} \wedge \phi^{j_2} \wedge \ldots \wedge \phi^{j_k} \not\equiv 0$$

because a_i can always be matched with ϕ^i through a permutation which only affects the sign. Consequently, $(a_i \lrcorner \omega) \equiv 0$ if and only if the coefficients d_J of all the terms containing ϕ^i are zero. □

Definition 4.3.10 *Let $\omega \in \Lambda^k(V^*)$. The associated space of ω is defined as*

$$A_\omega := \{v \in V | v \lrcorner \omega \equiv 0\}$$

The dual associated space of ω is defined as $A_\omega^\perp$.

Recall that the divisor space L_ω of a k-form ω contains all the 1-forms which can be factored from every term of ω. The dual associated space $A_\omega^\perp$ contains all the 1-forms which are contained in at least one term of ω. Therefore, $L_\omega \subset A_\omega^\perp$. The following result ties these notions together.

Lemma 4.3.14 *The following statements are equivalent:*

1. *A k-form $\omega \in \Lambda^k(V^*)$ is decomposable.*
2. *The divisor space L_ω has dimension k.*
3. *The dual associated space $A_\omega^\perp$ has dimension k.*
4. *$L_\omega = A_\omega^\perp$.*

Proof:

(1) ⇔ (2). If ω is decomposable, then there exists a set of basis vectors $\phi^1, \phi^2, \ldots, \phi^n$ for V^* such that

$$\omega = \phi^1 \wedge \ldots \wedge \phi^k.$$

Therefore, $L_\omega = span\{\phi^1, \phi^2, \ldots, \phi^k\}$ which has dimension k. Conversely, if L_ω has dimension k, then k terms can be factored from ω. Since ω is k-tensor, it must be decomposable.

(1) ⇔ (3). Let $a_1, \ldots, a_n$ be the basis of V which is dual to $\phi^1, \phi^2, \ldots, \phi^n$. Since

$$\omega = \phi^1 \wedge \ldots \wedge \phi^k,$$

ω is not a function of $a_{k+1}, \ldots, a_n$. Therefore,

$$A_\omega = span\{a_{k+1}, \ldots, a_n\}.$$

This implies that $A_\omega^\perp$ has dimension k. Conversely, if $A_\omega^\perp$ has dimension k, then A_ω has dimension $n-k$ which means that ω is a k-form which is a function of k variables. Therefore, it must have the form

$$\omega = \phi^1 \wedge \ldots \wedge \phi^k$$

for some linearly independent set of 1-forms $\phi^1, \phi^2, \ldots, \phi^k \in V^*$.

(2)&(3) ⇔ (4). It is always true that $L_\omega \subset A_\omega^\perp$. Therefore, if $dim(L_\omega) = dim(A_\omega^\perp)$, then $L_\omega = A_\omega^\perp$. It is also always true that $0 \leq dim(L_\omega) \leq k$ and $k \leq dim(A_\omega^\perp) \leq n$. Therefore, $L_\omega = A_\omega^\perp$ implies that $dim(L_\omega) = dim(A_\omega^\perp) = k$. □

4.4 The Pull Back of a Linear Transformation

The pull back of a linear transformation is a generalization of the dual map which we introduced in section 4.1.

Let T be a linear map from a vector space V to a vector space W, and let $f \in \mathcal{L}^k(W)$. Using T and f, a new multilinear function on V can be defined by mapping each k-tuple of vectors $v_1, \ldots, v_k \in V^k$ to $T(v_1), \ldots, T(v_k) \in W^k$ and then apply the multilinear function f. More formally,

Definition 4.4.1 *Let $T : V \to W$ be a linear transformation. The* _dual_ *or* _pull back transformation_

$$T^* : \mathcal{L}^k(W) \to \mathcal{L}^k(V)$$

is defined for all $f \in \mathcal{L}^k(W)$ by

$$(T^* f)(v_1, \ldots, v_k) := f(T(v_1), \ldots, T(v_k)).$$

Note that $T^* f$ is multilinear since T is a linear transformation. The pull back map T^* also has the following properties.

Lemma 4.4.1 *Let $T : V \to W$ be a linear transformation, and let*

$$T^* : \mathcal{L}^k(W) \to \mathcal{L}^k(V)$$

be the dual transformation. Then

1. *T^* is linear.*
2. *$T^*(f \otimes g) = T^* f \otimes T^* g$.*
3. *If $S : W \to X$ is linear, then $(S \circ T)^* f = T^*(S^* f)$.*

Proof: See Munkres [5] page 225. □

The following lemma says that the subspace of k-forms in $\mathcal{L}^k(V)$ is invariant under the action of a pull back mapping.

Lemma 4.4.2 *Let $T : V \longrightarrow W$ be a linear transformation. If f is an alternating tensor on W, then $T^* f$ is also an alternating tensor on V, and*

$$T^*(f \wedge g) = T^* f \wedge T^* g.$$

Proof: See Munkres [5] page 234. □

4.5 Contravariant Tensors

Up to this point, all the tensors which we have worked with have been defined as multilinear functions over the vector space V. If we simply replace V with V^* in all our definitions, then nothing is changed, and we can define an identical set of tensors over V^*.

A multilinear function $T : (V^*)^k \to \mathcal{R}$ is said to be a _contravariant tensor of_ _order k_. The set of all k-tensors on V^* is denoted by $\mathcal{L}^k(V^*)$ or $\underbrace{V \otimes \ldots \otimes V}_{k-times}$. Note that in this notation we are implicitly using the natural identification between V^{**} and V. For $k = 1$, we have $\mathcal{L}^k(V^*) = V$, i.e., the vector space itself. For this reason, contravariant tensors are sometimes called multivectors.

Chapter 5

Grassmann Manifolds

This chapter introduces a class of geometric objects called Grassmann manifolds. These manifolds are another basic building block for the bundles which will be introduced in Chapter 7. The chapter is divided into four sections. The first section defines a Grassmann manifold and discusses its construction and topology. The second section introduces a set of standard coordinate charts which provide a convenient local description of a Grassmann manifold. The third section discusses the mappings between Grassmann manifolds which are induced by linear transforms of $\mathcal{R}^n$. The projectivized linear group is introduced in this section, and some of its properties are discussed. The fourth section discusses the notion of an interval defined by two nested subspaces. For a more in-depth treatment of projective geometry, the interested reader is referred to the book by Mihalek [25].

5.1 The Topology of a Grassmann Manifold

Throughout this chapter, we will work with the set of all k-dimensional subspaces of $\mathcal{R}^n$ which we will denote by S_k^n. We will endow this set with a topology which turns it into an $(n-k) \times k$ dimensional manifold called the Grassmann Manifold of k-planes in $\mathcal{R}^n$. This manifold will be denoted by G_k^n.

We will construct a topology which turns S_k^n into a smooth manifold. We begin this construction with the vector space $\mathcal{R}^{n \times k}$ endowed with its standard topology. A point $M \in \mathcal{R}^{n \times k}$ can be regarded as an $n \times k$ matrix. If this matrix has full rank, then its columns form a set of linearly independent vectors which span a k-dimensional subspace V of $\mathcal{R}^n$. We will use $\mathcal{Q} \subset \mathcal{R}^{n \times k}$ to denote the set of all full rank matrices in $\mathcal{R}^{n \times k}$.

Two matrices $M, \hat{M} \in \mathcal{Q}$ span the same subspace V if and only if $M = \hat{M}T$ for some nonsingular $k \times k$ matrix T. Therefore, we can form a set of equivalence classes in $\mathcal{Q}$ with $M \sim \hat{M}$ if and only if there exists a nonsingular matrix $T \in \mathcal{R}^{k \times k}$ such that $M = \hat{M}T$. We will use $[M]$ to denote the equivalence class which contains the matrix M and $\mathcal{Q}/\sim$ to denote the set of all such equivalence classes. Since each matrix in the equivalence class maps to the same subspace, there is a bijection $\Psi : \mathcal{Q}/\sim \longrightarrow S_k^n$ between the set of equivalence classes and S_k^n. There is also a quotient map $\Phi : \mathcal{Q} \to \mathcal{Q}/\sim : M \to [M]$ which maps each element of $\mathcal{Q}$ to its corresponding equivalence class.

We are now ready to form a topology on S_k^n. We will proceed by first forming a topology on $\mathcal{Q}/\sim$ and then mapping this topology to S_k^n through the bijection Ψ. We define a subset of $U \subset \mathcal{Q}/\sim$ to be open if and only if $\Phi^{-1}(U)$ is open in $\mathcal{Q}$. The collection of all such subsets will be denoted by $\mathcal{T}$. The topology on S_k^n is defined to be the collection of open subsets

$$\hat{\mathcal{T}} = \{U \subset S_k^n \mid U = \Psi(V)\ V \in \mathcal{T}\}.$$

Endowed with these topologies, S_k^n and $\mathcal{Q}/\sim$ are homeomorphic topological spaces, so we can treat them as essentially the same object, and from this point on, we will not distinguish between the two. We can now formally define G_k^n to be the topological space $(S_k^n, \mathcal{T})$.

5.2 Local Coordinate Charts on G_k^n

Up to this point, we have shown that G_k^n is a topological space, but we still need to show that it is a manifold. To do this, we need to verify that every point has a neighborhood which is homeomorphic to $\mathcal{R}^s$ for some finite integer s.

Associated with any $n \times k$ matrix, we have $\binom{n}{k}$ $k \times k$ minors. Let $i \in 1, \ldots, \binom{n}{k}$ be an index set for these minors, and define M_i to be the $k \times k$ matrix corresponding to the ith minor. Let $\mathcal{Q}_i \subset \mathcal{Q}$ denote the dense open subset of $\mathcal{Q}$ consisting of all $M \in \mathcal{Q}$ such that the M_i is nonsingular. We can define a continuous onto map $\Phi_i : \mathcal{Q}_i \rightarrow \mathcal{R}^{(n-k)\times k}$ by setting $\Phi(M)$ equal to the rows of MM_i^{-1} which are not in the ith minor. The ith minor of MM_i^{-1} will always be the identity, so this mapping is completely determined by its other $(n-k) \times k$ elements. Furthermore, if $M \in \mathcal{Q}_i$, then so is MT for every nonsingular $T \in \mathcal{R}^{k\times k}$, and the mapping Φ_i maps the matrix MT to the rows of $MT(M_iT)^{-1}$ which are not in the ith minor. Simplifying this expression, we find that

$$MT(M_iT)^{-1} = MTT^{-1}M_i^{-1} = MM_i^{-1},$$

so the function Φ_i maps any two matrices in the same equivalence class to the same point. Therefore, Φ_i induces a mapping $\hat{\Phi}_i : \mathcal{Q}_i/\sim \rightarrow \mathcal{R}^{(n-k)\times k} : [M] \rightarrow \Phi_i(M)$ which is a homeomorphism.

Any matrix $M \in \mathcal{Q}$ is nonsingular, so it must contain at least one full rank minor; therefore, it is contained in some $\mathcal{Q}_i$. Furthermore, the fact that each element in the equivalence class $[M]$ is also contained in the same $\mathcal{Q}_i$ implies that each equivalence class $[M] \in \mathcal{Q}/\sim$ is contained in some open set $\mathcal{Q}_i/\sim$ which is homeomorphic to $\mathcal{R}^{(n-k)\times k}$. Consequently, G_k^n forms an $(n-k) \times k$-dimensional manifold, and the local coordinate charts $(\mathcal{Q}_i, \hat{\Phi}_i)$ form a C^∞ (actually analytic) coordinate atlas for G_k^n.

Example: The Manifold $\mathbf{G_2^3}$. As a specific example, we will consider the $(3-2) \times 2 = 2$ dimensional manifold G_2^3. The space $\mathcal{Q}^{3\times 2}$ is equal to the set of all 3×2 full-rank matrices, and it forms a dense open subset of $\mathcal{R}^{3\times 2}$. Let

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{bmatrix}$$

be any element of $\mathcal{Q}^{3\times 2}$. It has $\binom{3}{2} = 3$ minors

$$M_1 = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

$$M_2 = \begin{bmatrix} m_{11} & m_{12} \\ m_{31} & m_{32} \end{bmatrix}$$

$$M_3 = \begin{bmatrix} m_{21} & m_{22} \\ m_{31} & m_{32} \end{bmatrix}$$

The matrices MM_i^{-1} are given by

$$MM_1^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{m_{31}m_{22}-m_{32}m_{21}}{m_{11}m_{22}-m_{21}m_{12}} & \frac{-m_{31}m_{12}+m_{32}m_{11}}{m_{11}m_{22}-m_{21}m_{12}} \end{bmatrix}$$

$$MM_2^{-1} = \begin{bmatrix} 1 & 0 \\ \frac{m_{21}m_{32}-m_{22}m_{31}}{m_{11}m_{32}-m_{31}m_{12}} & \frac{-m_{21}m_{12}+m_{22}m_{11}}{m_{11}m_{32}-m_{31}m_{12}} \\ 0 & 1 \end{bmatrix}$$

$$MM_3^{-1} = \begin{bmatrix} \frac{m_{11}m_{32}-m_{12}m_{31}}{m_{21}m_{32}-m_{31}m_{22}} & \frac{-m_{11}m_{22}+m_{12}m_{21}}{m_{21}m_{32}-m_{31}m_{22}} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The space $\mathcal{Q}_1^{3\times 2}$ is equal to the dense open subset of $\mathcal{Q}^{3\times 2}$ where $m_{11}m_{22} - m_{21}m_{12} \neq 0$. Similarly, the spaces $\mathcal{Q}_2^{3\times 2}$ and $\mathcal{Q}_3^{3\times 2}$ are equal to dense open subsets of $\mathcal{Q}^{3\times 2}$ where $m_{11}m_{32} - m_{31}m_{12} \neq 0$ and $m_{21}m_{32} - m_{31}m_{22} \neq 0$, respectively. The maps Φ_i are defined by

$$\Phi_1 : M \to \begin{bmatrix} \frac{m_{31}m_{22}-m_{32}m_{21}}{m_{11}m_{22}-m_{21}m_{12}} & \frac{-m_{31}m_{12}+m_{32}m_{11}}{m_{11}m_{22}-m_{21}m_{12}} \end{bmatrix}$$

$$\Phi_2 : M \to \begin{bmatrix} \frac{m_{21}m_{32}-m_{22}m_{31}}{m_{11}m_{32}-m_{31}m_{12}} & \frac{-m_{21}m_{12}+m_{22}m_{11}}{m_{11}m_{32}-m_{31}m_{12}} \end{bmatrix}$$

$$\Phi_3 : M \to \begin{bmatrix} \frac{m_{11}m_{32}-m_{12}m_{31}}{m_{21}m_{32}-m_{31}m_{22}} & \frac{-m_{11}m_{22}+m_{12}m_{21}}{m_{21}m_{32}-m_{31}m_{22}} \end{bmatrix}.$$

Finally, the induced maps $\hat{\Phi}_i$ are defined by

$$\hat{\Phi}_1 : \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a_{31} & a_{32} \end{bmatrix} \right] \to \begin{bmatrix} a_{31} & a_{32} \end{bmatrix}$$

$$\hat{\Phi}_2 : \left[\begin{bmatrix} 1 & 0 \\ a_{21} & a_{22} \\ 0 & 1 \end{bmatrix} \right] \to \begin{bmatrix} a_{21} & a_{22} \end{bmatrix}$$

$$\hat{\Phi}_3 : \left[\begin{bmatrix} a_{11} & a_{12} \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right] \to \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}.$$

◇

5.3 Mappings Between Grassmann Manifolds

A linear automorphism $A : \mathcal{R}^n \to \mathcal{R}^n$ induces a unique diffeomorphism $\Phi(A) : G_k^n \to G_k^n : [M] \to [AM]$ which makes the following diagram commute

$$\begin{array}{ccc} \mathcal{Q}^{n\times k} & \xrightarrow{A} & \mathcal{Q}^{n\times k} \\ \Phi \downarrow & & \Phi \downarrow \\ G_k^n & \xrightarrow{\Phi(A)} & G_k^n \end{array}$$

The set of all such induced maps together with the composition operation forms a group which is called the <u>projective general linear group</u> on $\mathcal{R}^n$ and is denoted by $PGL(\mathcal{R}^n)$. To verify that this is actually a group, we note that

1. For any $\Phi(A), \Phi(B), \Phi(C) \in PGL(\mathcal{R}^n)$, we have that

$$\Phi(A) \circ (\Phi(B) \circ \Phi(C)) = (\Phi(A) \circ \Phi(B)) \circ \Phi(C)$$

since for all $[M] \in G_k^n$,

$$\begin{aligned} \Phi(A) \circ (\Phi(B) \circ \Phi(C))([M]) &= [A(BC)M] \\ &= [(AB)CM] \\ &= (\Phi(A) \circ \Phi(B)) \circ \Phi(C)([M]). \end{aligned}$$

2. There is a neutral element $\Phi(I) \in PGL(\mathcal{R}^n)$ which satisfies

$$\Phi(I) \circ \Phi(A) = \Phi(A) \circ \Phi(I) = \Phi(A)$$

for all $A \in GL(\mathcal{R}^n)$.

3. For any $A \in GL(\mathcal{R}^n)$, $\Phi(A)$ has an inverse $\Phi^{-1}(A) = \Phi(A^{-1})$.

Furthermore, the mapping Φ is a homomorphism from $GL(\mathcal{R}^n)$ onto $PGL(\mathcal{R}^n)$ since $(\Phi(A) \circ \Phi(B))([M]) = [ABM] = \Phi(AB)([M])$. The kernel of the homomorphism Φ is the set of all $A \in GL(\mathcal{R}^n)$ which satisfy $\Phi(A) = \Phi(I)$. This set is given by

$$\ker(\Phi) = \{A \in GL(\mathcal{R}^n) \mid A = \lambda \cdot I,\ \lambda \in \mathcal{R}\}$$

which is a closed subgroup of $GL(\mathcal{R}^n)$. A standard result from group theory says that any homomorphism $\phi : G \to H$ which maps a group G onto a group H induces a group isomorphism from $G/\ker(\phi)$ to H. So, in this instance, we have that $GL(\mathcal{R}^n)/\ker(\Phi)$ is isomorphic to $PGL(\mathcal{R}^n)$ where $GL(\mathcal{R}^n)/\ker(\Phi)$ is the set of equivalence classes corresponding to the equivalence relation

$$A \sim B \Leftrightarrow A = \lambda B, \quad \lambda \in \mathcal{R}.$$

A subspace $\Delta \subset \mathcal{R}^n$ which satisfies the equation $A(\Delta) \subset \Delta$ for a given automorphism $A \in GL(\mathcal{R}^n)$ is said to be an invariant subspace of A. It is easy to see that if Δ_k is a k-dimensional invariant subspace of A, then it is also a fixed point of the induced map $\Phi(A) : G_k^n \to G_k^n$. We can characterize the set of all k-dimensional invariant subspaces of A. To do this, we will work on the coordinate chart of G_k^n defined by the matrix

$$\Delta = \begin{bmatrix} m_1^1 & \cdots & & m_k^1 \\ \vdots & & & \vdots \\ m_1^{n-k} & \cdots & & m_k^{n-k} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ & & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The action of the automorphism A on Δ can be represented by the matrix multiplication

$$\begin{bmatrix} A_1^1 & A_2^1 \\ A_1^2 & A_2^2 \end{bmatrix} \begin{bmatrix} M \\ I \end{bmatrix} = \begin{bmatrix} A_1^1 M + A_2^1 \\ A_1^2 M + A_2^2 \end{bmatrix} \sim \begin{bmatrix} (A_1^1 M + A_2^1)(A_1^2 M + A_2^2)^{-1} \\ I \end{bmatrix}.$$

From this equation, it is clear that Δ is a fixed point of $\Phi(A)$ if and only if M is a solution to the equation

$$\begin{aligned} (A_1^1 M + A_2^1)(A_1^2 M + A_2^2)^{-1} &= M \\ \Leftrightarrow M A_1^2 M + M A_2^2 &= A_1^1 M + A_2^1 \\ \Leftrightarrow M A_1^2 M + M A_2^2 - A_1^1 M - A_2^1 &= 0. \end{aligned}$$

5.4 Intervals

Given two subspaces Δ_1 and Δ_2 which satisfy the condition $\Delta_1 \subset \Delta_2$, their interval is defined to be the collection of all subspaces D which satisfy

$$\Delta_1 \subset D \subset \Delta_2. \tag{5.1}$$

We will denote this collection by $S^{[\Delta_1,\Delta_2]}$ and the set of all k-dimensional subspaces which satisfy 5.1 by $S_k^{[\Delta_1,\Delta_2]}$. Since $S_k^{[\Delta_1,\Delta_2]} \subset S_k^n$, the topology of G_k^n induces a subspace topology for $S_k^{[\Delta_1,\Delta_2]}$ which turns it into a smooth submanifold of G_k^n. We will use $G_k^{[\Delta_1,\Delta_2]}$ to denote this submanifold.

Lemma 5.4.1 *Suppose that Δ_1 and Δ_2 are respectively s-dimensional and $s+t$-dimensional subspaces of $\mathcal{R}^n$, then the manifold $G_{s+k}^{[\Delta_1,\Delta_2]}$ is diffeomorphic to the Grassmann manifold G_k^t for every $0 \le k \le t$.*

Proof: To prove this lemma, we need to show that there exists a linear map T taking $\mathcal{R}^n$ onto $\mathcal{R}^t$ which satisfies

$$\begin{aligned} T(\Delta_1) &= 0 \\ T(\Delta_2) &= \mathcal{R}^t. \end{aligned} \tag{5.2}$$

If such a mapping exists, it will induce the desired diffeomorphism $\Phi(T) : G_{s+k}^{[\Delta_1,\Delta_2]} \to G_k^t$. To show that such a mapping exists, we form a basis of $\mathcal{R}^n$ using the columns of a matrix of the form

$$\begin{bmatrix} I & V_2^1 & V_3^1 \\ 0 & I & V_3^2 \\ 0 & 0 & I \end{bmatrix}$$

where the right block of columns span Δ_1 and the middle and right blocks of columns span Δ_2. The inverse of this matrix is

$$\begin{bmatrix} I & -V_2^1 & (V_2^1 V_3^2 - V_3^1) \\ 0 & I & -V_3^2 \\ 0 & 0 & I \end{bmatrix}$$

which we will view as a mapping taking $\mathcal{R}^n \to \mathcal{R}^n$. Selecting the middle row of this matrix, we obtain the matrix

$$\begin{bmatrix} 0 & I & -V_3^2 \end{bmatrix}$$

which corresponds to a mapping taking $\mathcal{R}^n \to \mathcal{R}^t$. Furthermore, this mapping clearly satisfies the conditions 5.2. □

As a consequence of lemma 5.4.1, if $A : \mathcal{R}^n \to \mathcal{R}^m$, $m < n$ is a surjective linear map, it induces a diffeomorphism between the manifold G_k^m and the manifold $G_{(n-m+k)}^{[\ker(A),\mathcal{R}^n]}$ consisting of all $(k+n-m)$-dimensional subspaces of $\mathcal{R}^n$ which contain the $(n-m)$-dimensional subspace $\ker(A)$. The homeomorphism $\Phi(A) : G_{(n-m+k)}^{[\ker(A),\mathcal{R}^n]} \to G_k^m$ is defined pointwise by the equations

$$\Delta \in G_k^m,\ \Delta \to A^{-1}(\Delta) = \Phi^{-1}(A)(\Delta)$$

$$\Gamma \in G_{(n-m+k)}^{[\ker(A),\mathcal{R}^n]}\ \Gamma \to A(\Gamma) = \Phi(A)(\Gamma).$$

Similarly, if $A : \mathcal{R}^m \to \mathcal{R}^n$, $m < n$ is an injective linear map, it induces a diffeomorphism between the manifold G_k^m and the manifold $G_k^{[0,Im(A)]}$ consisting of all k-dimensional subspaces of $\mathcal{R}^n$ which are contained in the m-dimensional subspace Im(A). Finally, if $A : \mathcal{R}^n \to \mathcal{R}^m$ is an arbitrary linear map with a kernel of dimension s, then it induces a diffeomorphism between the manifold $G_k^{[0,Im(A)]}$ and the manifold $G_{(k+s)}^{[\ker(A),\mathcal{R}^n]}$.

Chapter 6

Exterior Algebra and Systems of Exterior Equations

This chapter discusses exterior algebra and systems of exterior equations. In section 4.3.3 of Chapter 4, we introduced the wedge product and interior product and discussed some of their properties. In the first section of this chapter, we will look more closely at the algebraic properties these operations give to the space of all alternating tensors. In the second section of this chapter, we introduce systems of exterior equations. This subject involves both the exterior algebra and Grassmann manifolds, and will serve as a prototype for the discussion of exterior differential systems in Chapter 8.

6.1 The Exterior Algebra of a Vector Space

In this section, we will consider the direct sum of the space of all 0-forms, 1-forms, 2-forms, etc.

$$\Lambda(V^*) = \Lambda^0(V^*) \oplus \Lambda^1(V^*) \oplus \Lambda^2(V^*) \oplus \cdots \oplus \Lambda^n(V^*).$$

This is clearly a vector space, and the wedge product acts as a multiplication operator on it, so it satisfies Definition B.0.1 of an algebra given in Appendix B. The pair $(\Lambda(V^*), \wedge)$ is called the <u>exterior algebra</u> over V^*. In this notation, any $\xi \in \Lambda(V^*)$ may be written as

$$\xi = \xi_0 + \xi_1 + \ldots + \xi_n$$

where each $\xi_p \in \Lambda^p(V^*)$.

Since $(\Lambda(V^*), \wedge)$ has the unity element $1 \in \Lambda^0(V^*)$, Theorem B.0.4 implies that the ideal generated by a finite set of elements $\Sigma := \{\alpha^i \in \Lambda(V^*), 1 \leq i \leq K\}$ can be written as

$$I_S = \{\pi \in \Lambda(V^*) |\, \pi = \sum_{i=1}^{K} \theta^i \wedge \alpha^i, \theta^i \in \Lambda(V^*)\}$$

6.2 Systems of Exterior Equations

In this section, we are going to use the exterior algebra to study a system of equations of the form

$$\alpha^1 = 0, \ldots, \alpha^K = 0$$

where each $\alpha^i \in \Lambda(V^*)$. The first thing we need to clarify is exactly what a "solution" to these equations means.

Definition 6.2.1 *A system of exterior equations on V is a finite set of linearly independent equations*

$$\alpha^1 = 0, \ldots, \alpha^K = 0$$

where each $\alpha^i \in \Lambda^k(V^*)$ for some $1 \leq k \leq n$. A solution to a system of exterior equations is any subspace $W \subset V$ such that

$$\alpha^1|_W \equiv 0, \ldots, \alpha^K|_W \equiv 0$$

where $\alpha|_W$ means that the arguments of $\alpha(v_1, \ldots, v_k)$ satisfy $v_1, \ldots, v_k \in W$.

A system of exterior equations generally does not have a unique solution since any subspace $W_1 \subset W$ will satisfy $\alpha|_{W_1} \equiv 0$ if $\alpha|_W \equiv 0$. In fact, rather than focusing on a particular solution, we will often use a system of exterior equations to define the submanifold of a Grassmann bundle of k-planes consisting of all points in the Grassmann bundle which solve the system of exterior equations.

A central fact concerning systems of exterior equations is given by the following lemma.

Lemma 6.2.1 *Given a system of exterior equations*

$$\alpha^1 = 0, \ldots, \alpha^K = 0 \tag{6.1}$$

and the corresponding ideal I_Σ generated by the collection of alternating tensors

$$\Sigma := \{\alpha^1, \ldots, \alpha^K\}, \tag{6.2}$$

a subspace W solves the system of exterior equations if and only if it also satisfies $\pi|_W \equiv 0$ for every $\pi \in I_A$.

Proof: If $\pi \in I_A$, then $\pi = \sum_{i=1}^K \theta^i \wedge \alpha^i$. Furthermore, if W satisfies $\pi|_W = 0$, then

$$\pi|_W = (\sum_{i=1}^K \theta^i \wedge \alpha^i)|_W \equiv 0$$

for some set of $\theta^i \in \Lambda(V^*)$. Since this equation must hold for every $\pi \in I_A$ and the α^i are assumed to be linearly independent, it implies

$$\alpha^1|_W \equiv 0, \ldots, \alpha^K|_W \equiv 0. \tag{6.3}$$

Conversely, if equation 6.3 holds, then $\pi|_W \equiv 0$ for all $\pi \in I_A$. □

Recall that an algebraic ideal was defined in a coordinate free way as a subspace of the algebra satisfying certain closure properties. Thus, the ideal has an intrinsic geometric meaning, and we can think of two sets of generators as representing the same system of exterior equations if they generate the same algebraic ideal.

Definition 6.2.2 *Two sets of generators, Σ_1 and Σ_2, are said to be algebraically equivalent if and only if they generate the same ideal, i.e., $I_{\Sigma_1} = I_{\Sigma_2}$.*

6.2.1 The Associated and Retracting Spaces

We will exploit the notion of equivalence to represent a system of exterior equations in a simplified form. The main tools we need in order to do this are called the associated space and retracting space of the system of exterior equations.

Definition 6.2.3 *Let Σ be a system of exterior equations and I_Σ the ideal which it generates. The associated space of the ideal I_Σ is defined as*

$$A(I_\Sigma) := \{v \in V | v \lrcorner \alpha \in I_\Sigma \ \forall \ \alpha \in I_\Sigma\}.$$

The dual associated space or retracting space of the ideal is defined as $A(I_\Sigma)^\perp$ and denoted by $C(I_\Sigma)$.

Once we have determined $C(I_\Sigma)$, we can find an algebraically equivalent system Σ' which is a subset of $\Lambda(C(I_\Sigma))$.

Lemma 6.2.2 *Let Σ be a system of exterior equations and I_Σ its corresponding algebraic ideal. Then there exists an algebraically equivalent system Σ' such that $\Sigma' \subset \Lambda(C(I_\Sigma))$.*

Proof: Let $v_1, \ldots, v_n$ be a basis for V and $\phi^1, \ldots, \phi^n$ be the corresponding dual basis. Assume that the basis has been selected such that the vectors $v_{r+1}, \ldots, v_n$ span $A(I_\Sigma)$. Consequently, the covectors $\phi^1, \ldots, \phi^r$ must span $C(I_\Sigma)$.

Let α be any k-form in Σ. Consider the form

$$\alpha' = \alpha - \phi^{r+1} \wedge (v_{r+1} \lrcorner \alpha).$$

Taking the interior product of α' with v_{r+1} gives,

$$v_{r+1} \lrcorner \alpha' = v_{r+1} \lrcorner \alpha - v_{r+1} \lrcorner \alpha + \phi^{r+1} \wedge (v_{r+1} \lrcorner (v_{r+1} \lrcorner \alpha)) \equiv 0.$$

Therefore, using lemma 4.3.13, we can conclude that α' has no terms involving ϕ^{r+1}. Since $v_{r+1} \in A(I_\Sigma)$, we know that $v_{r+1} \lrcorner \alpha \in I$. Therefore, we can replace α with α' in the set of generators, and the ideal generated will be unchanged since

$$\theta \wedge \alpha = \theta \wedge \alpha' + \theta \wedge \phi^{r+1} \wedge (v_{r+1} \lrcorner \alpha)$$

$$\Rightarrow \theta \wedge \alpha = \theta \wedge \alpha' \mod I.$$

We can continue this process for $v_{r+2}, \ldots, v_n$ to produce a k-form $\hat{\alpha}$ which is generated by elements of $\Lambda(C(I_\Sigma))$. □

The following example is taken from Yang [7].

Example: Let $v_1, \ldots, v_6$ be a basis for $\mathcal{R}^6$, and let $\theta^1, \ldots, \theta^6$ be the dual basis. Consider the system of exterior equations

$$\alpha^1 = \theta^1 \wedge \theta^3 = 0,$$

$$\alpha^2 = \theta^1 \wedge \theta^4 = 0,$$

$$\alpha^3 = \theta^1 \wedge \theta^2 - \theta^3 \wedge \theta^4 = 0,$$

$$\alpha^4 = \theta^1 \wedge \theta^2 \wedge \theta^5 - \theta^3 \wedge \theta^4 \wedge \theta^6 = 0.$$

The set of generators Σ is given by

$$\Sigma = \{\alpha^1, \alpha^2, \alpha^3, \alpha^4\},$$

and the ideal I_Σ is given by

$$I_\Sigma := \{\xi \in \Lambda(V^*) \mid \xi = \sum_{i=1}^{4} \pi^i \wedge \alpha^i,\ \pi^i \in \Lambda(V^*)\}.$$

The associated space of I_Σ is defined by

$$A(I_\Sigma) := \{v \in \mathcal{R}^6 \mid v \lrcorner \pi \in I_\Sigma \ \forall\, \pi \in I_\Sigma\}.$$

Because I_Σ contains no 1-forms, we can infer that

$$v \lrcorner \alpha^1 = 0,\ v \lrcorner \alpha^2 = 0,\ \text{and}\ v \lrcorner \alpha^3 = 0,\ \forall\ v \in A(I_\Sigma).$$

Expanding the first equation, we get

$$v \lrcorner \alpha^1 = v \lrcorner (\theta^1 \wedge \theta^3) = (v \lrcorner \theta^1) \wedge \theta^3 + (-1)^1 \theta^1 \wedge (v \lrcorner \theta^3) = \theta^1(v)\theta^3 - \theta^3(v)\theta^1 = 0$$

which implies that $\theta^1(v) = 0$ and $\theta^3(v) = 0$. Similarly,

$$v \lrcorner \alpha^2 = \theta^1(v)\theta^4 - \theta^4(v)\theta^1 = 0$$

$$v \lrcorner \alpha^3 = \theta^1(v)\theta^2 - \theta^2(v)\theta^1 - \theta^3(v)\theta^4 + \theta^4(v)\theta^3 = 0$$

implying that $\theta^2(v) = 0$ and $\theta^4(v) = 0$. Therefore, we can conclude that

$$A(I_\Sigma) \subset span\{v_5, v_6\}.$$

Evaluating the equation $v \lrcorner \alpha^4$ gives

$$\begin{aligned} v \lrcorner \alpha^4 &= (v \lrcorner (\theta^1 \wedge \theta^2)) \wedge \theta^5 + (-1)^2 (\theta^1 \wedge \theta^2) \wedge (v \lrcorner \theta^5) \\ &\quad - (v \lrcorner (\theta^3 \wedge \theta^4)) \wedge \theta^6 - (-1)^2 (\theta^3 \wedge \theta^4) \wedge (v \lrcorner \theta^6) \\ &= \theta^5(v)\theta^1 \wedge \theta^2 - \theta^6(v)\theta^3 \wedge \theta^4 \\ &= a(\theta^1 \wedge \theta^3) + b(\theta^1 \wedge \theta^4) + c(\theta^1 \wedge \theta^2 - \theta^3 \wedge \theta^4). \end{aligned}$$

Equating coefficients, we find that

$$\theta^5(v) = \theta^6(v) = c, \ \forall\, v \in A(I_\Sigma).$$

Now v must be of the form $v = xv_5 + yv_6$, so we get

$$\theta^5(xv_5 + yv_6) = x = c$$

$$\theta^6(xv_5 + yv_6) = y = c.$$

Therefore, $A(I_\Sigma) = span\{(v_5 + v_6)\}$. If we select as a new basis for $\mathcal{R}^6$ the vectors

$$w_i = v_i, \ i = 1, \ldots, 4, \ w_5 = v_5 - v_6, \ w_6 = v_5 + v_6.$$

then the new dual basis becomes

$$\gamma^i = \theta^i, \ i = 1, \ldots, 4, \ \gamma^5 = \frac{\theta^5 - \theta^6}{2}, \ \gamma^6 = \frac{\theta^5 + \theta^6}{2}.$$

With respect to this new basis, the retracting space $C(I_\Sigma)$ is given by

$$C(I_\Sigma) = span\{\gamma^1, \ldots, \gamma^5\}.$$

In these coordinates, the generator set becomes

$$\Sigma' = \{\gamma^1 \wedge \gamma^3, \ \gamma^1 \wedge \gamma^4, \ \gamma^1 \wedge \gamma^2 - \gamma^3 \wedge \gamma^4, \ \gamma^1 \wedge \gamma^2 \wedge \gamma^5\} \subset \Lambda(C(I_\Sigma)).$$

◇

The following lemma allows us to find the dimension of the retracting space in the special case where the generators of the ideal are a collection of 1-forms together with a single 2-form.

Lemma 6.2.3 *Let I_Σ be an ideal generated by the set*

$$\Sigma = \{\omega^1, \ldots, \omega^s, \Omega\}$$

where $\omega^i \in V^*$ and $\Omega \in \Lambda^2(V^*)$. Let r be the smallest integer such that

$$(\Omega)^{r+1} \wedge \omega^1 \wedge \ldots \wedge \omega^s = 0.$$

Then the retracting space $C(I_\Sigma)$ is of dimension $2r + s$.

Proof: See Bryant [8] pages 11-12. □

6.2.2 Independence Conditions

A system of exterior equations with independence condition Ω is a pair (Σ, Ω) where Σ is a set of generators and Ω is a decomposable k-form. A solution to the system (Σ, Ω) is any subspace $\Delta \subset V$ which satisfies the conditions

1. For each $\alpha \in \Sigma$, $\alpha|_\Delta = 0$.
2. $\Omega|_\Delta \neq 0$.

The second condition implies that every solution must be at least k-dimensional, and also restricts the set of k-dimensional solutions to lie in a local coordinate chart of the Grassmann manifold G^n_k. The following example illustrates these points.

Example: Let $(\Sigma, \psi^1 \wedge \cdots \wedge \psi^m)$ be a system of exterior equations with an independence condition. Let $\phi^1, \ldots, \phi^{(n-m)}$ be any collection of one-forms which, taken together with the ψ's, forms a basis for V^*.

$$\mathcal{R}^{n*} = \text{span}\{\phi^1, \ldots, \phi^{(n-m)}, \psi^1, \ldots, \psi^m\} \tag{6.4}$$

We can also form the basis for V which is dual to the cobasis 6.4.

$$\mathcal{R}^n = \text{span}\{e_1, \ldots, e_{(n-m)}, v_1, \ldots, v_m\} \tag{6.5}$$

so that

$$\begin{aligned} e_i \lrcorner \phi^j &= \delta^j_i \\ e_i \lrcorner \psi^j &= 0 \\ v_i \lrcorner \psi^j &= \delta^j_i \\ v_i \lrcorner \phi^j &= 0. \end{aligned}$$

Any element $\Delta \in G^n_m$ which satisfies the independence condition

$$\psi^1 \wedge \cdots \wedge \psi^m|_\Delta \neq 0$$

can be represented with respect to the basis 6.5 as the matrix

$$\begin{bmatrix} M \\ I \end{bmatrix}$$

where the elements m^i_j form a local coordinate chart for the Grassmann bundle. We can form a new basis for V using the columns of the matrix

$$\begin{bmatrix} I & M \\ 0 & I \end{bmatrix} \tag{6.6}$$

defined relative to the old basis 6.5. The inverse of this matrix is

$$\begin{bmatrix} I & -M \\ 0 & I \end{bmatrix}$$

and its rows form a new cobasis which is dual to the basis 6.6. We will define

$$\omega^i = \phi^i - \sum_{j=1}^{m} m^i_j \psi^j \tag{6.7}$$

so that this new cobasis can be described by

$$\mathcal{R}^{n*} = \text{span}\{\omega^1, \ldots, \omega^{(n-m)}, \psi^1, \ldots, \psi^m\}. \tag{6.8}$$

Suppose that $\alpha^k \in \Sigma$ is a one-form. With respect to the basis 6.5, this form can be written as

$$\alpha^k = \sum_{j=1}^{(n-m)} a_j^k \phi^j + \sum_{l=1}^{m} b_l^k \psi^l.$$

With respect to the cobasis 6.8, this can be written as

$$\alpha^k = \sum_{j=1}^{(n-m)} a_j^k \omega^j + \sum_{i=1}^{m} (\sum_{j=1}^{(n-m)} a_j^k m_i^j + b_i^k) \psi^i.$$

Therefore, the subspace $[M] \in G_m^n$ satisfies the one-forms in Σ if and only if

$$\sum_{j=1}^{(n-m)} a_j^k m_i^j + b_i^k = 0$$

for each one-form $\alpha^k \in \Sigma$. In matrix form, this looks like

$$A^k M = B^k,$$

so the set of all solutions forms an affine subset of $\mathcal{R}^{(n-m)\times m}$.

Next, we can consider a two form $\beta^k \in \Sigma$

$$\beta^k = \sum_{r=1}^{(n-m)} \sum_{t=1}^{(n-m)} \frac{1}{2} a_{rt}^k \phi^r \wedge \phi^t + \sum_{p=1}^{(n-m)} \sum_{q=1}^{m} b_{pq}^k \phi^p \wedge \psi^q + \sum_{v=1}^{m} \sum_{w=1}^{m} \frac{1}{2} c_{vw}^k \psi^v \wedge \psi^w.$$

Written with respect to the new basis, this equation becomes

$$\begin{aligned} \beta^k \quad &= \quad \sum_{1 \le l < j \le m} \left[\sum_{r=1}^{(n-m)} \sum_{t=1}^{(n-m)} a_{rt}^k m_j^r m_l^t + \sum_{p=1}^{(n-m)} (b_{pj}^k m_l^p - b_{pl}^k m_j^p) + c_{lj}^k \right] \psi^l \wedge \psi^j \\ &\mod \{\omega^1, \ldots, \omega^{(n-m)}\}. \end{aligned}$$

Restricted to the plane $[M]$, this form will be identically zero if and only if the m_j^i satisfy the quadratic equations

$$\sum_{r=1}^{(n-m)} \sum_{t=1}^{(n-m)} a_{rt}^k m_j^r m_l^t + \sum_{p=1}^{(n-m)} (b_{pj}^k m_l^p - b_{pl}^k m_j^p) + c_{lj}^k = 0$$

which can be written in matrix form as

$$M^t A^k M + M^t B^k - (B^k)^T M + C^k = 0$$

with $A^k = -(A^k)^T$ and $C^k = -(C^k)^T$. Therefore, if a system of exterior equations consists of 1-forms and 2-forms, then the subset of points in G_m^n which are solutions will generally be described by a collection of linear and quadratic algebraic equations. $\diamond$

Chapter 7

Bundle Structures

In this chapter, we are going to extend the objects which have been defined in the previous three chapters with respect to a vector space to analogous objects which are defined with respect to the tangent bundle of a smooth manifold. In order to do this in a rigorous fashion, we need to introduce a class of objects called fibre bundles. The first part of this chapter discusses the general definition of a fibre bundle, as well as the definitions of and constructions for several special classes of fibre bundles which are analogous to tensor spaces and Grassmann manifolds. The second part of this chapter discusses the concept of a section of a fibre bundle. The presentation focuses on the sections of the fibre bundles which are described in the first section, and discusses some of the algebraic operations which are associated with these sections. A more comprehensive treatment of this material can be found in the book by Husemoller [15].

7.1 Fibre Bundles

A locally-finite fibre bundle consists of a sextuple $(E, B, \pi, V, G, \mathcal{A})$ where

1. E, B, and V are topological spaces called the total space, base space, and standard fibre of the bundle, respectively.

2. π is a continuous map taking E onto B in such a way that the inverse image of each point $p \in B$, $\pi^{-1}(p)$, is homeomorphic to V. The set $\pi^{-1}(p)$ is called the fibre over p.

3. The space G is a topological group which has an associated left action on V.

4. $\mathcal{A}$ is a maximal atlas of charts. A chart is a pair (U, Ψ) where $U \subset B$ is an open subset of B, and $\Psi : \pi^{-1}(U) \to U \times V$ is a homeomorphism taking fibres of the bundle over U to the product space $U \times V$. The homeomorphism must induce the identity map on U, so for any $p \in U$, $\pi^{-1}(p) \to \{p\} \times V$. An atlas of charts is any collection of charts (U_i, Ψ_i) defined such that the U_i form an open cover of B and such that if (U_i, Ψ_i) and (U_j, Ψ_j) are two charts which both contain a point p, then $(\Psi_i \circ \Psi_j^{-1})(p) : \{p\} \times V \to \{p\} \times V$ is an element of G. An atlas can be enlarged by adding other charts which satisfy these compatibility requirements. In particular, there exists a unique maximal enlargement of any atlas consisting of all additional charts which satisfy the compatibility requirements.

Although B need only be a topological space, we will only be concerned with fibre bundles for which the base space is a smooth manifold.

7.1.1 The Tangent Bundle of a Smooth Manifold

In Appendix A, the tangent bundle is defined as the union over M of the set of all point derivations at each point $p \in M$. Given a smooth manifold M with a coordinate atlas $\mathcal{A}_M$, we can construct a fibre bundle with base space M and group $GL(\mathcal{R}^n)$ which is isomorphic to the tangent bundle which is defined in Appendix A. To do this, we will follow a standard construction which, with slight variations, will be used to produce all the bundles discussed in this chapter. We will use $\mathcal{A}_M$ as an index set and form the disjoint union

$$N = \cup_{(U_x,x)\in\mathcal{A}_M} U_x \times \mathcal{R}^n.$$

If this set is given the maximal topology such that the inclusion maps

$$i_x : U_x \times \mathcal{R}^n \to N$$

are all continuous, it becomes a smooth manifold. Each point in this manifold is a triple of the form $((U_x, x), p, v)$ where $(U_x, x) \in \mathcal{A}_M$, $p \in U_x$, and $v \in \mathcal{R}^n$. We will define an equivalence relation on this manifold by declaring two points $((U_x, x), p, v)$ and $((U_y, y), q, w)$ to be equivalent if and only if

$$\begin{aligned} p &= q \\ w &= \left.\frac{\partial y}{\partial x}\right|_p v. \end{aligned} \tag{7.1}$$

We will denote the equivalence class of $((U_x, x), p, v)$ by $[((U_x, x), p, v)]$. We will let E denote the set of all such equivalence classes, and endow E with the quotient topology in the standard fashion to turn it into a topological space. We can define a projection π from E onto M by

$$\pi([((U_x, x), p, v)]) \to p$$

and an atlas of charts of the form (U_x, Ψ_x) by

$$\Psi_x : \pi^{-1}(U_x) \to U_x \times \mathcal{R}^n \ : \ [((U_x, x), p, v)] \to (p, v).$$

Using the equivalent relation 7.1, we find that for any pair of charts (U_x, Ψ_x) and (U_y, Ψ_y), the mapping

$$\Psi_y \circ \Psi_x^{-1} : (U_x \cap U_y) \times \mathcal{R}^n \to (U_x \cap U_y) \times \mathcal{R}^n$$

can be expressed pointwise as

$$\Psi_y \circ \Psi_x^{-1} : (p, v) \to (p, \frac{\partial y}{\partial x}|_{x^{-1}(p)} v),$$

so that

$$\Psi_y \circ \Psi_x^{-1}|_{\{p\}\times\mathcal{R}^n} \in GL(\mathcal{R}^n).$$

This collection of charts can then be extended to a maximal atlas which we denote by $\mathcal{A}$. The sextuple

$$TM := (E, B, \pi, \mathcal{R}^n, GL(\mathcal{R}^n), \mathcal{A})$$

is the desired fibre bundle. If one works through a few computations in local coordinates, it is not difficult to see how this definition corresponds to the definition of the tangent bundle given in Appendix A. However, a formal proof that these bundles are isomorphic is rather long and will not be included here. The interested reader is referred to Spivak [6].

7.1.2 Tensor Bundles

Once we have defined the tangent bundle as in the previous section, we can, with only slight modification, produce a bundle of tensors of any order over M. We will denote this bundle by

$$\mathcal{L}^k(M) := (E, B, \pi, \mathcal{L}^k(\mathcal{R}^n), GL(\mathcal{R}^n), \mathcal{A})$$

where $T \in GL(\mathcal{R}^n)$ acts on $\mathcal{L}^k(\mathcal{R}^n)$ through the pull back map

$$T^* : \mathcal{L}^k(\mathcal{R}^n) \to \mathcal{L}^k(\mathcal{R}^n) \ : \ \omega \to T^*\omega.$$

As before, we will use $\mathcal{A}_M$ as an index set and form the disjoint union

$$N := \cup_{(U_x,x)\in\mathcal{A}_M} U_x \times \mathcal{L}^k(\mathcal{R}^n).$$

Endowed with the maximal topology such that each of the inclusion maps

$$i_x : U_x \times \mathcal{L}^k(\mathcal{R}^n) \to N$$

is continuous, N becomes a topological space. We define an equivalence relation on N by declaring any two points $((U_x,x),p,\omega)$ and $((U_y,y),q,\theta)$ to be equivalent if and only if

$$\begin{aligned} p &= q. \\ \omega &= \left.\frac{\partial y}{\partial x}^*\right|_p \theta. \end{aligned} \qquad (7.2)$$

If we denote the equivalence class corresponding to $((U_x,x),p,\omega)$ by $[((U_x,x),p,\omega)]$ and denote the topological space corresponding to the set of all such equivalence classes equipped with the quotient topology as E, then we can define a continuous surjection $\pi : [((U_x,x),p,\omega)] \to p$ and local trivializations

$$\Psi_x : \pi^{-1}(U_x) \to U_x \times \mathcal{L}^k(\mathcal{R}^n) \ : \ [((U_x,x),p,\omega)] \to (p,\omega)$$

which satisfy

$$(\Psi_y \circ \Psi_x^{-1})|\{p\} \times \mathcal{L}^k(\mathcal{R}^n) = \left.\left(\frac{\partial y}{\partial x}^{-1}\right)^*\right|_p .$$

7.1.3 Partial Frame and Coframe Bundles

In the construction for the tangent bundle, we can replace $\mathcal{R}^n$ with $\mathcal{R}^{n\times m}$, or equivalently $\mathcal{R}^n \otimes (\mathcal{R}^m)^*$, and replace the equivalence relation with the equivalence relation

$$((U_x,x),p,M) \sim ((U_y,y),q,\hat{M}) \ \Leftrightarrow$$

$$\begin{aligned} p &= q \\ \hat{M} &= \left.\frac{\partial y}{\partial x}\right|_p M \end{aligned} \qquad (7.3)$$

where $\hat{M}, M \in \mathcal{R}^{n\times m}$. A point $[((U_x,x),p,M)]$ corresponds to a selection of m tangent vectors at the point p. Note that this is also equivalent to the condition that the two points $((U_x,x),p,v\otimes\theta)$ and $((U_y,y),q,w\otimes\gamma)$ satisfy

$$\begin{aligned} p &= q \\ \theta &= \gamma \\ w\otimes\gamma &= \left(\left.\frac{\partial y}{\partial x}\right|_p v\right)\otimes\theta. \end{aligned} \qquad (7.4)$$

Similarly, in the construction of the tensor bundle $\mathcal{L}^k(\mathcal{R}^n)$, we can replace each standard fibre with the space $\mathcal{R}^m \otimes \mathcal{L}^k(\mathcal{R}^n)$ and the equivalence relation 7.2 with the equivalence relation

$$\begin{aligned} p &= q \\ w &= v \\ w \otimes \gamma &= v \otimes \left(\left.\frac{\partial y}{\partial x}\right|_p^* \theta \right) \end{aligned} \tag{7.5}$$

where $v, w \in \mathcal{R}^m$ and $\gamma, \theta \in \mathcal{L}^k(\mathcal{R}^n)$. Any element of $\mathcal{R}^m \otimes \mathcal{L}^k(\mathcal{R}^n)$ is called a vector-valued k-tensor, and the corresponding bundle is called a vector-valued k-tensor bundle. A point $[((U_x, x), p, w \otimes \omega)]$ in this bundle corresponds to the choice of m different k-tensors at the point p.

In both of these constructions, it is sometimes useful to consider only collections of vectors or k-tensors which are linearly independent. If $m \leq n$, the set of all elements of $\mathcal{R}^n \otimes (\mathcal{R}^m)^*$ in which the columns are linearly independent forms a dense open subset of $\mathcal{R}^n \otimes (\mathcal{R}^m)^*$. We will denote this set by $\mathcal{Q} \subset \mathcal{R}^n \otimes (\mathcal{R}^m)^*$. In the construction just outlined, we can replace the standard fibre $\mathcal{R}^n \otimes (\mathcal{R}^m)^*$ with $\mathcal{Q}$ and use the same equivalence relation. The resulting bundle is called an m-frame bundle over M. Similarly, if $m \leq dim(\mathcal{L}^k(\mathcal{R}^n))$, then the set of points in the space $\mathcal{R}^m \otimes \mathcal{L}^k(\mathcal{R}^n)$ at which the collection of tensors is linearly independent forms a dense open subset which we can use as the standard fibre for a bundle.

7.1.4 Grassmann Bundles of k-planes in $\mathcal{R}^n$

In this section, we will once again repeat the standard construction, but this time, we will use the Grassmann manifold G_k^n as the standard fibre.

Suppose that we are given an n-dimensional manifold M together with an atlas of coordinate charts (U_j, x_j). We will form a new manifold N which consists of the disjoint union of the manifolds $U_j \times G_k^n$. One can think of this as making a separate copy of the domain of each coordinate chart, and then forming the product space with the Grassmann manifold of k-planes in $\mathcal{R}^n$. We endow N with the largest topology such that each inclusion map $i_j : U_j \times G_k^n \rightarrow N$ is continuous. Each point of the manifold M has the form (x_j, p, Δ) where x_j specifies the coordinate chart, $p \in U_j$ specifies a point in the domain of x_j, and $\Delta \in G_k^n$. We next form an equivalence class on N using the equivalence relation $(x_j, p, \Delta) \sim (x_r, q, \Gamma)$ if and only if

1. $p = q$
2. $\Delta = \Phi\left(\frac{\partial x_j}{\partial x_r}|_p\right)(\Gamma)$ where Φ is the mapping which takes the Jacobian matrix to its corresponding element in $PGL(\mathcal{R}^n)$.

We will denote the equivalence class of (x_j, p, Δ) by $[x_j, p, \Delta]$. The set of such equivalence classes, endowed with the quotient topology, will form the total space E of the Grassmann Bundle. The projection π is defined pointwise by $\pi([x_j, p, \Delta]) = p$. Associated with each U_j, we can define a mapping $\Phi_j : \pi^{-1}(U_j) \rightarrow U_j \times G_k^n$ by the pointwise assignment $\Phi_j([x_j, p, \Delta]) = (p, \Delta)$, so that the pair (U_j, Φ_j) forms a chart on the bundle. Finally, for any two charts $(U_1, \Phi_1), (U_2, \Phi_2)$, we have that at any point $p \in U_1 \cap U_2$,

$$(\Phi_2 \circ \Phi_1^{-1})(p, \Delta) = \Phi_2([x_1, p, \Delta]) =$$

$$\Phi_2\left(\left[x_2, p, \Phi\left(\left.\frac{\partial x_j}{\partial x_r}\right|_p \Delta\right)\right]\right) = \left(p, \Phi\left(\left.\frac{\partial x_j}{\partial x_r}\right|_p \Delta\right)\right).$$

The U_j form an open cover of M, and the charts satisfy the compatibility conditions prescribed in the definition of a Grassmann bundle, so they form an atlas for the bundle.

7.2 Sections of Fibre Bundles

A section of a fibre bundle $(E, B, \pi, V, G, \mathcal{A})$ is a function $s : B \to E$ defined such that $\pi(s(p)) = p$ for each $p \in B$. In other words, the function maps each point p to an element in the fibre $\pi^{-1}(\{p\})$ over p. The section is continuous if s is a continuous function. If E and B are both smooth manifolds, then the section is said to be smooth if s is a smooth function. A local section over $U \subset M$ is a function $s_u : U \to \pi^{-1}(U)$ which satisfies the condition that $\pi(s_u(p)) = p$ for each $p \in U$.

A vector field is an important example of a bundle section. Other examples of bundle sections include covector fields, tensor fields, distributions, and codistributions. In this section, we will discuss each of these types of bundle sections and will develop some algebraic operations defined in relation to them.

7.2.1 Vector Fields

A vector field $X : M \to TM$ is a section of the tangent bundle TM. If X is of class C^∞, it is called a smooth section of TM. Recall that with respect to a coordinate chart (U, x) containing a point $p \in M$, a tangent vector can be expressed as

$$X_p = \sum_{i=1}^{n} a_i \frac{\partial}{\partial x^i}.$$

Similarly, a vector field X can be locally expressed as

$$X(p) = \sum_{i=1}^{n} a^i(p) \frac{\partial}{\partial x^i}.$$

From this equation, it is clear that the vector field X is C^∞ if and only if the scalar functions $a^i : M \longrightarrow \mathcal{R}$ are C^∞.

We will let $V(M)$ denote the collection of all smooth sections of the tangent bundle. We can give $V(M)$ two different algebraic structures. First, we can turn $V(M)$ into an infinite dimensional vector space over $\mathcal{R}$. Vector addition on this space is defined pointwise, so for every $X_1, X_2 \in V(M)$,

$$(X_1 + X_2)(p) := X_1(p) + X_2(p),$$

and scalar multiplication is defined pointwise by $(aX)(p) := a \cdot X(p)$ for every $a \in \mathcal{R}$. Second, we can give $V(M)$ the structure of a module over the ring of smooth functions on M, $C^\infty(M)$. In this case, addition is defined pointwise as above, and scalar multiplication is defined pointwise by $aX := a(p) \cdot X(p)$ for every $a \in C^\infty(M)$.

A vector field provides a geometric description of a differential equation. An integral curve of a vector field X is a mapping $c : (-\varepsilon, \varepsilon) \longrightarrow M$ whose tangent vector at each point is identically equal to the vector field at that point. The theory of ordinary differential equations guarantees that every smooth vector field determines a unique integral curve passing through each point $p \in M$. Depending on the vector field, the parameter ε may be limited, or we may be able to expand the interval $(-\varepsilon, \varepsilon)$ to be the whole real line.

Given a function $h : M \longrightarrow \mathcal{R}$, we will often be interested in the rate at which h changes along an integral curve of a vector field. This rate of change is called the Lie derivative of h along the vector field X. The Lie derivative is denoted as $L_X h$ and is formally defined by the equation

$$L_X h = X(h).$$

We can also define a product operation on $V(M)$ called the Lie bracket. Given two vector fields X and Y, their Lie bracket is denoted by $[X, Y]$ and is defined to be the unique vector field which satisfies the equation

$$[X, Y](h) := X(Y(h)) - Y(X(h)).$$

for every $h \in C^\infty(p)$. In particular, if we choose the coordinate functions x^i, we get

$$[X,Y](x^i) = [X,Y]_i = \sum_j \frac{\partial Y_i}{\partial x^j} X_j - \sum_j \frac{\partial X_i}{\partial x^j} Y_j$$

and we therefore obtain

$$[X,Y](x) = \frac{\partial Y}{\partial x} X(x) - \frac{\partial X}{\partial x} Y(x)$$

The Lie bracket is skew-symmetric

$$[X,Y] = -[Y,X]$$

and also satisfies the Jacobi identity

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0.$$

The Lie bracket and the Lie derivative of a smooth function are related by the following equation

$$[aX, bY] = ab[X,Y] + a(L_X b)Y - b(L_Y a)X \tag{7.6}$$

for every $X, Y \in V(M)$ and every $a, b \in C^\infty(M)$.

7.2.2 Tensor Fields

A k-tensor field ω is a section of the bundle $\mathcal{L}^k(M)$

$$\omega : M \to \mathcal{L}^k(M).$$

At each point $p \in M$, $\omega(p)$ defines a multilinear function mapping k-tuples of tangent vectors in T_pM to $\mathcal{R}$. That is

$$\omega(p) : \underbrace{T_pM \times \cdots \times T_pM}_{k\ \ times} \to \mathcal{R}.$$

In particular, if ω is a section of $\Lambda^k(M)$, then ω is called a differential form of order k or differential k-form on M. In this case, $\omega(p)$ defines an alternating k-tensor at each point $p \in M$. We will denote the collection of all sections of the bundle $\Lambda^k(M)$ by $\Omega^k(M)$, and the space of all sections of the bundle $\Lambda(M)$ will be denoted by

$$\Omega(M) := \Omega^0(M) \oplus \ldots \oplus \Omega^n(M).$$

At each point $p \in M$, let $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ denote standard basis for T_pM, and let the 1-forms ϕ^i denote the dual basis defined such that

$$\phi^i(p)(\frac{\partial}{\partial x^j}) = \delta_{ij}.$$

Recall that the set of k-tensors defined by

$$\phi^I = \phi^{i_1} \otimes \phi^{i_2} \otimes \ldots \otimes \phi^{i_k}$$

for each multi-index $I = (i_1, \ldots, i_k)$ form a basis for $\mathcal{L}^k(T_pM)$ and that the set of all alternating k-forms defined by

$$\psi^I = \phi^{i_1} \wedge \phi^{i_2} \wedge \ldots \wedge \phi^{i_k}$$

for each ascending multi-index $I = (i_1, \ldots, i_k)$ form a basis for $\Lambda^k(T_pM)$. A k-tensor ω on M can be uniquely written as

$$\omega(p) = \sum_I b_I(p)\phi^I(p)$$

for multi-index I and scalar functions $b_I(p)$. Likewise, a k-form α can be written uniquely as

$$\alpha(p) = \sum_I c_I(p)\psi^I(p)$$

for ascending multi-index I and scalar functions c_I. The k-tensor ω and k-form α are of class C^∞ if and only if the functions b_I and c_I are of class C^∞, respectively. Given two forms $\omega \in \Omega^k(M), \theta \in \Omega^l(M)$, we have

$$\omega = \sum_I b_I\psi^I$$

$$\theta = \sum_I c_I\psi^I$$

$$\omega \wedge \theta = \sum_I \sum_J b_I c_I \psi^I \wedge \psi^J.$$

Recall that we have defined $\Lambda^0(T_pM) = \mathcal{R}$. As a result, the space of differential forms of order 0 on M is simply the space of all functions $f : M \longrightarrow \mathcal{R}$, and the wedge product of $f \in \Omega^0(M)$ and $\omega \in \Omega^k(M)$ is defined as

$$(w \wedge f)(p) = (f \wedge w)(p) = f(p) \cdot w(p).$$

7.2.3 The Exterior Derivative

The <u>differential df</u> of a 0-form f is defined pointwise as the unique 1-form which satisfies the equation

$$df(p)(X_p) = X_p(f)$$

for every $X_p \in T_pM$. The operator d is linear on 0-forms; that is,

$$d(af + bg) = a \cdot df + b \cdot dg.$$

This follows from the fact that X_p is a linear operator.

Using this operator d, we obtain a new way of expressing the elementary 1-forms $\phi^i(p)$ on T_pM. Let $x : M \longrightarrow \mathcal{R}^n$ be a coordinate function in a neighborhood of p, and consider the differentials of the coordinate functions

$$dx^i(p)(X_p) = X_p(x^i).$$

If we evaluate the differentials dx^i at the basis tangent vectors of T_pM, we obtain

$$dx^i(p)(\frac{\partial}{\partial x^j}) = \delta_{ij}$$

and therefore the operator d maps each function x^i to its corresponding element in the dual basis. In short, $dx^i(p) = \phi^i(p)$. Consequently, the differentials $dx^i(p)$ form a basis for $\mathcal{L}^1(T_pM)$ and any k-tensor ω can be uniquely written as

$$\omega(p) = \sum_I b_I(p)dx^I(p)$$

for multi-index I. Similarly, any k-form can be uniquely written as

$$\omega(p) = \sum_I b_I(p)dx^I(p)$$

for ascending multi-index I. Therefore, any k-tensor ω can be expressed in the chart (U, x) containing p as

$$\omega(p) = \sum_{i=1}^{n} b_I(p) dx^1 \otimes \cdots \otimes dx^n$$

while the k-form α is expressed as

$$\alpha(p) = \sum_{i=1}^{n} c_I(p) dx^1 \wedge \ldots \wedge dx^n.$$

Using this basis, we find that the operator d takes any 0-form $f \in C^\infty(M)$ to the 1-form

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i.$$

We now have an operator d which takes 0-forms to 1-forms. We would like to extend this operator to all of $\Omega(M)$. In order to accomplish this, we will inductively define an operator $d : \Omega^k(M) \longrightarrow \Omega^{k+1}(M)$. We will then prove that this extension is unique in the sense that there is only one operator d which is compatible with the operator which takes 0-forms to 1-forms and which satisfies the properties listed in Theorem 7.2.1 below.

Definition 7.2.1 *Let ω be a k-form on a manifold M which can be described with respect to the coordinate chart (U, x) by the equation*

$$\omega = \sum_I a_I dx^I$$

for ascending multi-index I. The exterior derivative or differential operator, d, is a linear map taking the k-form ω to the (k+1)-form $d\omega$ by

$$d\omega = \sum_I da_I \wedge dx^I.$$

Notice that each a_I is a smooth function whose differential da_I is defined by

$$da_I = \sum_{j=1}^{n} \frac{\partial a_I}{\partial x^j} dx^j.$$

Consequently, we find that for any k-form,

$$d\omega = \sum_I \sum_{j=1}^{n} \frac{\partial a_I}{\partial x^j} dx^j \wedge dx^I.$$

One can see from the definition that this operator is certainly linear. The next theorem precisely states the sense in which this operator is also unique.

Theorem 7.2.1 *Let M be a manifold and let $p \in M$. Then the exterior derivative is the unique linear operator*

$$d : \Omega^k(M) \to \Omega^{k+1}(M)$$

for $k \geq 0$ that satisfies

1. *If f is a 0-form, then df is the 1-form*

$$df(p)(X_p) = X_p(f)$$

2. *If $\omega^1 \in \Omega^k(M), \omega^2 \in \Omega^l(M)$, then*

$$d(\omega^1 \wedge \omega^2) = d\omega^1 \wedge \omega^2 + (-1)^k \omega^1 \wedge d\omega^2$$

3. *For every form ω, $d(d\omega) = 0$.*

Proof: Property (1) can be easily checked from the definition of the exterior derivative. We now prove property (2). Because of linearity of the exterior derivative it suffices to consider the case $\omega^1 = f dx^I$ and $\omega^2 = g dx^J$ in some chart (U, x). We have

$$\begin{aligned} d(\omega^1 \wedge \omega^2) &= d(fg) \wedge dx^I \wedge dx^J \\ &= gdf \wedge dx^I \wedge dx^J + fdg \wedge dx^I \wedge dx^J \\ &= d\omega^1 \wedge \omega^2 + (-1)^k f dx^I \wedge dg \wedge dx^J \\ &= d\omega^1 \wedge \omega^2 + (-1)^k \omega^1 \wedge d\omega^2 \end{aligned}$$

We now prove property (3). Again it suffices to consider the case $\omega = f dx^I$ because of linearity. Since f is a 0-form,

$$d(df) = d(\sum_{j=1}^{n}(D_j f)dx^j) = \sum_{i=1}^{n}\sum_{j=1}^{n} D_i(D_j f)dx^i \wedge dx^j$$

$$= (D_i(D_j f) - D_j(D_i f))dx^i \wedge dx^j = 0$$

where $D_i f$ is the standard derivative $\frac{\partial f}{\partial x^i}$. If $\omega = f dx^I$ is a k-form, then $d\omega = df \wedge dx^I$, and since

$$d(dx^I) = d(1 \wedge dx^I) = d(1) \wedge dx^I = 0,$$

we get

$$d(d\omega) = d(df) \wedge dx^2 - df \wedge d(dx^I) = 0.$$

We now show that d is the unique linear operator with the above properties. Assume that d' is another linear operator with the same properties. Consider again a k-form $\omega = f dx^I$. Since d' satisfies property (2), we have

$$d'(f dx^I) = d'f \wedge dx^I + f \wedge d'(dx^I).$$

From the above formula, we see that if we can show that $d'(dx^I) = 0$, then we will get

$$d'(f dx^I) = d'f \wedge dx^I = d(f dx^I)$$

which will complete the proof. We therefore need to show that

$$d'(dx^1 \wedge \ldots \wedge dx^k) = 0. \tag{7.7}$$

Both d and d' satisfy property (1), so we must have

$$dx^I = dx^{i_1} \wedge \ldots \wedge dx^{i_k} = d'x^{i_1} \wedge \ldots \wedge d'x^{i_k} = d'x^I$$

since the coordinate functions x^i are 0-forms.

We prove equation 7.7 by induction. It can be easily checked to hold for $k = 0$. Assume that equation 7.7 holds for $k-1$. Then define

$$\eta = dx^2 \wedge \ldots \wedge dx^k.$$

Then

$$d'(dx^I) = d'(d'x^{i_1} \wedge d'x^{i_k}) = d'(d'x^{i_1}) \wedge \eta - d'x_{i_1} \wedge d'\eta = 0$$

since d' also satisfies property (3) and $d'\eta = d\eta$ by the induction hypothesis. □

Now let $f : M \longrightarrow N$ be a smooth map between two manifolds. We have seen that the push forward map, f_*, is a linear transformation from T_pM to $T_{f(p)}N$. Therefore, given tensors or forms on $T_{f(p)}N$, we can use the pull back tranformation, f^*, in order to define tensors or forms on T_pM. The next theorem shows that the exterior derivative and the pull back transformation commute.

Theorem 7.2.2 *Let $f : M \longrightarrow N$ be a smooth map between manifolds. If ω is a k-form on N, then*

$$f^*(d\omega) = d(f^*\omega)$$

Proof: See Spivak [6] pages 295-296. □

We can define the interior product of a tensor field and a vector field pointwise as the interior product of a tensor and a tangent vector.

Definition 7.2.2 *Given a k-form $\omega \in \Omega^k(M)$ and a vector field X, the interior product or anti-derivation of ω with X is a $(k-1)$ form defined pointwise by*

$$(X(p) \lrcorner \omega(p))(v_1, \ldots, v_{k-1}) = \omega(p)(X(p), v_1, \ldots, v_{k-1}).$$

Therefore, an antiderivation of a k-form ω simply substitutes the first argument with the given vector and thus results in a (k-1)-form.

The following lemma establishes a relation between the exterior derivative and Lie brackets.

Lemma 7.2.1 (Cartan's Formula) *Let $\omega \in \Omega^1(M)$ be a 1-form and $X, Y \in V(M)$ be smooth vector fields. Then*

$$\begin{aligned} d\omega(X,Y) &= X(\omega(Y) - Y(\omega(x)) - \omega([X,Y]) \\ &= X \lrcorner d(Y \lrcorner \omega) - Y \lrcorner d(X \lrcorner \omega) - [X,Y] \lrcorner \omega. \end{aligned}$$

Proof: Because of linearity, it is adequate to consider $\omega = fdg$. The left-hand side of the above formula is

$$\begin{aligned} d\omega(X,Y) &= df \wedge dg(X,Y) \\ &= df(X) \cdot dg(Y) - df(Y) \cdot dg(X) \\ &= X(f) \cdot Y(g) - Y(f) \cdot X(g) \end{aligned}$$

while the right-hand side is

$$\begin{aligned} X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]) &= X(fY(g)) - Y(fX(g)) - f(XY(g) - YX(g)) \\ &= X(f) \cdot Y(g) - Y(f) \cdot X(g) \end{aligned}$$

which completes the proof. □

The Lie derivative of a differential k-form $\omega \in \Omega^k(M)$ with respect to a vector field X is defined by the equation

$$L_X\omega := X \lrcorner d\omega + d(X \lrcorner \omega). \tag{7.8}$$

Lemma 7.2.2 *The operator $L_X(\cdot) : \Omega(M) \to \Omega(M)$ satisfies the following properties*

1. *For every $a, b \in C^\infty(M)$,*

$$L_{aX}(b\omega) = abL_X\omega + a(L_Xb)\omega + bda \wedge (X \lrcorner \omega)$$

2. $L_X(\omega^1 \wedge \omega^2) = L_X\omega^1 \wedge \omega^2 + \omega^1 \wedge L_X\omega^2$

3. $L_X d\omega = dL_X\omega$.

Proof: Each of these statements can be verified directly using equation 7.8 together with the properties of the exterior derivative and the interior product. □

Finally, we have the following result.

Lemma 7.2.3 *Let $v, r \in V(M)$ be two vector fields over M. The operator*

$$L_v(r \lrcorner (\cdot)) - r \lrcorner L_v(\cdot) : \Omega(M) \to \Omega(M) \tag{7.9}$$

is equal to the operator

$$[v,r] \lrcorner (\cdot) : \Omega(M) \to \Omega(M).$$

Proof: Let K denote the set of all ascending k-tuples in $\{1,\ldots,n\}$ and $\omega \in \Omega^k(M)$ be any k-form which can be described relative to the basis

$$\{dx^1,\ldots,dx^n\}$$

by the equation

$$\omega = \sum_{I\in K} a_I dx^I$$

where $I=\{i_1,\ldots,i_k\}$ and

$$dx^I = dx^{i_1}\wedge\cdots\wedge dx^{i_k}.$$

If we compute

$$L_v(r \lrcorner \omega) - r \lrcorner (L_v\omega),$$

we get

$$L_v(r \lrcorner \sum_{I\in K} a_I dx^I) - r \lrcorner L_v(\sum_{I\in K} a_I dx^I) =$$

$$\sum_{I\in K}\Big(L_v(a_I)(r\lrcorner dx^I) + a_I L_v(r\lrcorner dx^I) - L_v(a_I)(r\lrcorner dx^I) - a_I(r\lrcorner L_v dx^I)\Big)$$

$$= \sum_{I\in K} a_I\left(L_v(r\lrcorner dx^I) - r\lrcorner L_v dx^I\right).$$

We also have that

$$[v,r]\lrcorner(\sum_{I\in K} a_I dx^I) = \sum_{I\in K} a_I([v,r]\lrcorner dx^I),$$

so the operators are equivalent if they are equivalent on a set of basis elements dx^I.

If dx^i is any 1-form in the cobasis, then Cartan's formula gives

$$0 = v\lrcorner r\lrcorner d\circ dx^i = r\lrcorner d(v\lrcorner dx^i) - v\lrcorner d(r\lrcorner dx^i) + [v,r]\lrcorner dx^i.$$

Moving the first two terms on the right over to the left, we have

$$\begin{aligned} v\lrcorner d(r\lrcorner dx^i) - r\lrcorner d(v\lrcorner dx^i) &= [v,r]\lrcorner dx^i \\ \Rightarrow L_v(r\lrcorner dx^i) - r\lrcorner L_v dx^i &= [v,r]\lrcorner dx^i. \end{aligned}$$

So the operators are equivalent on the set of 1-forms. Suppose that the operators are equivalent on the set of k-forms. Let $I\in K$ be any ascending k-tuple. If dx^i is any basis element for which $i\notin I$, then the form $dx^i\wedge dx^I$ is a basis element for the $(k+1)$-forms. We first compute

$$L_v(r\lrcorner(dx^i\wedge dx^I)) = L_v((r\lrcorner dx^i)dx^I - dx^i\wedge(r\lrcorner dx^I))$$

$$= L_v(r\lrcorner dx^i)dx^I + (r\lrcorner dx^i)L_v dx^I - L_v dx^i\wedge(r\lrcorner dx^I) - dx^i\wedge L_v(r\lrcorner dx^I). \qquad (7.10)$$

Similarly, we can compute

$$r\lrcorner L_v(dx^i\wedge dx^I) = r\lrcorner((L_v dx^i)\wedge dx^I) + r\lrcorner(dx^i\wedge L_v dx^I) =$$

$$(r\lrcorner(L_v dx^i))dx^I - (L_v dx^i)\wedge(r\lrcorner dx^I) + (r\lrcorner dx^i)L_v dx^I - dx^i\wedge(r\lrcorner L_v dx^I). \qquad (7.11)$$

Finally, subtracting equation 7.11 from equation 7.10, we get

$$L_v(r\lrcorner(dx^i\wedge dx^I) - r\lrcorner L_v(dx^i\wedge dx^I) =$$

$$L_v(r\lrcorner dx^i)dx^I - dx^i\wedge L_v(r\lrcorner dx^I) - (r\lrcorner L_v dx^i))dx^I + dx^i\wedge(r\lrcorner L_v dx^I)$$

which can be rearranged to give

$$= [L_v(r \lrcorner\, dx^i) - (r \lrcorner\, L_v dx^i))]dx^I - dx^i \wedge [L_v(r \lrcorner\, dx^I) - (r \lrcorner\, L_v dx^I)].$$

Using the fact that the operators are equivalent on 1-forms together with the induction hypothesis, the above expression can be rewritten as

$$([v,r] \lrcorner\, dx^i)dx^I - dx^i \wedge ([v,r] \lrcorner\, dx^I) = [v,r] \lrcorner\, dx^i \wedge dx^I.$$

So the two operators are also equivalent on $(k+1)$-forms; and by induction, this must be true for all k. □

7.2.4 Distributions and Codistributions

A smooth k-dimensional distribution is a smooth section of the Grassmann bundle of k-planes over a manifold M. Similarly, a smooth k-dimensional codistribution is a smooth section of the Grassmann bundle $G_k^{T^*M}(M)$ whose fibres over a point $p \in M$ consist of all k-dimensional subspaces of T_p^*M.

In general, a Grassmann bundle may not have a smooth section. For example, consider the Grassmann bundle of one-dimensional subspaces over the two-sphere $G_1^2(S^2)$. This is a well-defined Grassmann bundle whose fibres above any point $p \in S^2$ consist of all the one-dimensional subspaces contained in T_pS^2, but the topology of S^2 does not admit a smooth mapping $f : S^2 \rightarrow G_1^2(S^2)$ which satisfies $\pi \circ f(p) = p$. However, locally a Grassmann bundle will always have a smooth section. Therefore, when working with distributions, we will often restrict our discussion to an open neighborhood $U \subset M$ over which a section exists. When dealing with the tangent bundle or a tensor bundle, we do not have to worry about the existence of a smooth section because the zero section always exists, but the topology of the base manifold may still place restrictions on the form of the sections. For example, the tangent bundle of the two-sphere has smooth sections, but every section must have at least one zero point.

In order to get around this complication, a distribution is often defined as the pointwise span of collection of smooth vector fields. Since the dimension of the subspace spanned by a collection of vector fields may change as the base point is varied, a distribution defined in this way need not correspond to the section of any Grassmann manifold. Any point at which the span of the vector fields is maximal is called a regular point of the distribution, and all other points are called singular points. There is a neighborhood around any regular point on which the pointwise span of the vector fields coincides with a local section of a Grassmann bundle.

A third way to define a distribution or codistribution is as $C^\infty(M)$ submodule of $V(M)$ or $\Omega^1(M)$, respectively. This is the definition which will be adopted for the remainder of this section. This definition is not quite equivalent to the second definition. In order to see this, consider the two vector fields $x\frac{\partial}{\partial x}$ and $x^2\frac{\partial}{\partial x}$ which are both defined over the manifold $\mathcal{R}$. Pointwise, these vector fields span the same subspace, so according to the second definition, they define the same distribution. However, they do not generate the same submodule of $V(\mathcal{R})$ since the vector field $x\frac{\partial}{\partial x} \notin \text{span}\,\{x^2\frac{\partial}{\partial x}\}$. If the dimension of the pointwise span of the submodule is constant on a neighborhood, $U \subset M$, then this definition is locally equivalent to the other two.

Given a collection of vector fields $v_1, \ldots, v_r$ defined over a smooth manifold M, their span is the $C^\infty(M)$ module defined by

$$\text{span}\{v_1, \ldots, v_r\} := \{v \in V(M) \mid v = a^i v_i \;\; a^i \in C^\infty(M)\}.$$

Similarly, given a collection of covector fields $\omega^1, \ldots, \omega^r$ defined over a smooth manifold M, their span is the $C^\infty(M)$ module defined by

$$\text{span}\{\omega^1, \ldots, \omega^r\} := \{\omega \in \Omega^1(M) \mid \omega = a_i\omega^i \;\; a_i \in C^\infty(M)\}.$$

For the remainder of this section we will define a smooth distribution to be any $C^\infty(M)$ submodule of $V(M)$ and a smooth codistribution to be any $C^\infty(M)$ submodule of $\Omega^1(M)$.

Given two smooth distributions Δ_1 and Δ_2, their sum will be denoted by $\Delta_1 + \Delta_2$ and defined by

$$\Delta_1 + \Delta_2 := \text{span}\{v \in V(M) \mid v = v_1 + v_2,\ \ v_1 \in \Delta_1 \text{ and } v_2 \in \Delta_2\}.$$

The smooth intersection of Δ_1 and Δ_2 will be denoted by $\Delta_1 \cap \Delta_2$ and defined by

$$\Delta_1 \cap \Delta_2 := \{v \in V(M) \mid v \in \Delta_1 \text{ and } v \in \Delta_2\}.$$

Similarly, given two smooth codistributions I_1 and I_2, their sum will be denoted by $I_1 + I_2$ and defined by

$$I_1 + I_2 := \text{span}\{\omega \in \Omega^1(M) \mid \omega \in I_1 \text{ or } \omega \in I_2\}.$$

The smooth intersection of I_1 and I_2 will be denoted by $I_1 \cap I_2$ and defined by

$$I_1 \cap I_2 := \{\omega \in \Omega^1(M) \mid \omega \in I_1 \text{ and } \omega \in I_2\}.$$

It is straightforward to verify that the smooth intersection of a distribution or codistribution is closed under modular addition and scalar multiplication; hence, it forms a submodule of $V(M)$ or $\Omega^1(M)$, respectively.

Finally, we will denote the smooth perp of a smooth distribution Δ (codistribution I) by $\Delta^\perp$ ($I^\perp$) and define it by

$$I^\perp := \{v \in V(M) \mid v \lrcorner \omega \equiv 0 \text{ for all } \omega \in I\}$$

$$\Delta^\perp := \{\omega \in \Omega^1(M) \mid v \lrcorner \omega \equiv 0 \text{ for all } v \in \Delta\}.$$

Given any two subspace V_1 and V_2 of a finite-dimensional vector space V, the following relationships hold

$$(V_1^\perp)^\perp = V_1 \tag{7.12}$$

$$(V_1 + V_2)^\perp = V_1^\perp \cap V_2^\perp \tag{7.13}$$

$$(V_1 \cap V_2)^\perp = V_1^\perp + V_2^\perp. \tag{7.14}$$

We want to investigate the extent to which similar relationship hold between the operations which we have defined for distributions and codistributions. Restricted to a point $p \in M$, a distribution "looks" like a subspace of a finite-dimensional vector space, and a codistribution "looks" like a subspace of the dual space to a finite-dimensional vector space, so we should expect similar relationships to hold. However, there are some subtleties involved which are related to the fact that we are working with $C^\infty(M)$-modules rather than vector spaces.

Lemma 7.2.4 *The following relationships hold between the sum, intersection, and perp of a smooth distribution or codistribution. Let D, D_1, and D_2 be smooth, finitely-generated distributions or codistributions.*

1. $D \subset (D^\perp)^\perp$.
2. *If D has constant dimension on an open set U, then, restricted to U, $D = (D^\perp)^\perp$.*
3. $(D_1 + D_2)^\perp = D_1^\perp \cap D_2^\perp$.
4. $D_1^\perp + D_2^\perp \subset (D_1 \cap D_2)^\perp$.
5. *If D_1, D_2, and $D_1 \cap D_2$ have constant dimension on an open set U, then, restricted to U, $(D_1 \cap D_2)^\perp = D_1^\perp + D_2^\perp$.*

Proof: The first statement follows directly from the definition of the perp,

$$(D^\perp)^\perp = \{v \in V(M) \mid v \lrcorner \omega \equiv 0 \; \forall \, \omega \in D^\perp\}.$$

Since $\omega \in D^\perp$ implies that $v \lrcorner \omega \equiv 0$ for all $v \in D$, it follows that $D \subset (D^\perp)^\perp$.

Assume that, restricted to U, D is a constant dimensional, finitely-generated distribution. Since D has constant dimension on U, for every point $p \in U$, there exists an open neighborhood U_p and a coordinate chart (x, U_p) such that the following conditions hold

1. The matrix of coefficients for the local representation of the generators can be written in the form
$$\begin{bmatrix} A_1(x) \\ A_2(x) \end{bmatrix}$$
where $A_2(x)$ has full rank.

2. Every element of D can be uniquely written as a linear combination of the columns of $A(x)$.

3. The span of the columns of $A(x)$ is equal to the span of the columns of
$$\begin{bmatrix} A_1(x)A_2^{-1}(x) \\ I \end{bmatrix}.$$

If we form the block-triangular matrix

$$B(x) = \begin{bmatrix} I & A_1(x) \\ 0 & A_2(x) \end{bmatrix},$$

then the columns of $B(x)$ span $V(U_p)$ and the rows of the inverse matrix

$$B^{-1}(x) = \begin{bmatrix} I & -A_1(x)A_2^{-1}(x) \\ 0 & A_2^{-1}(x) \end{bmatrix}$$

span $\Omega^1(U_p)$. Consequently, any $\omega \in \Omega^1(U_p)$ can be written uniquely as some linear combination

$$\omega(x) = \sum_{j=1}^{n-m} \sum_{i=1}^{n} c_j(x) b_i^j(x) dx^i + \sum_{k=n-m+1}^{n} c_k(x) dx^k.$$

If $\omega \in D^\perp$, then $v \lrcorner \omega \equiv 0$ for all $v \in D$. This can only happen if $c_k(x) \equiv 0$ for $k > n - m$. Consequently,

$$D^\perp = \text{span}\{\sum_{i=1}^{n} b_i^j(x) dx^i\}, \; j = 1, \ldots, n - m.$$

If we repeat the same argument starting with $D^\perp$, then we find that, restricted to U_p, $D = (D^\perp)^\perp$. This construction can be repeated in an open neighborhood of every point p, so it must hold true on all of U. Furthermore, if $v_1, \ldots, v_r$ are generators for D and $v_1, \ldots, v_n$ are generators for $\Omega^1(U)$, then $v_{r+1}^*, \ldots, v_n^*$ are generators for $D^\perp$.

To prove the second statement, we begin by noting that

$$\begin{aligned} D_1 &\subset D_1 + D_2 \\ D_2 &\subset D_1 + D_2 \end{aligned}$$

which implies that

$$\begin{aligned} (D_1 + D_2)^\perp &\subset D_1^\perp \\ (D_1 + D_2)^\perp &\subset D_2^\perp \end{aligned}$$

and

$$(D_1 + D_2)^\perp \quad \subset \quad D_2^\perp \cap D_1^\perp.$$

To prove equality, we also have to show that

$$D_2^\perp \cap D_1^\perp \quad \subset \quad (D_1 + D_2)^\perp.$$

If $\omega \in D_2^\perp \cap D_1^\perp$, then by definition, $v_1 \lrcorner\, \omega \equiv 0$ for all $v_1 \in D_1$ and $v_2 \lrcorner\, \omega \equiv 0$ for all $v_2 \in D_2$. Consequently,

$$(\alpha \cdot v_1 + \beta \cdot v_2) \lrcorner\, \omega \equiv 0$$

for arbitrary $\alpha, \beta \in C^\infty(M)$. Every vector

$$v \in D_1 + D_2$$

can be written as $v = \alpha \cdot v_1 + \beta \cdot v_2$ for some $\alpha, \beta \in C^\infty(M)$, $v_1 \in D_1$, and $v_2 \in D_2$. Therefore, $\omega \in (D_1 + D_2)^\perp$.

To prove the third statement, we follow a similar sequence

$$\begin{aligned} D_1 \cap D_2 \quad &\subset \quad D_1 \\ D_1 \cap D_2 \quad &\subset \quad D_2 \end{aligned}$$

which implies that

$$\begin{aligned} D_1^\perp \quad &\subset \quad (D_1 \cap D_2)^\perp \\ D_2^\perp \quad &\subset \quad (D_1 \cap D_2)^\perp \end{aligned}$$

and

$$D_1^\perp + D_2^\perp \quad \subset \quad (D_1 \cap D_2)^\perp$$

If D_1 and D_2 have constant dimensions on U, then $(D_1^\perp)^\perp = D_1$ and $(D_2^\perp)^\perp = D_2$. From the second statement, we have that

$$(D_1^\perp + D_2^\perp)^\perp = (D_1^\perp)^\perp \cap (D_2^\perp)^\perp$$

which implies

$$(D_1^\perp + D_2^\perp)^\perp \subset (D_1^\perp)^\perp \cap (D_2^\perp)^\perp.$$

The equality of the distributions with their double perps implies

$$(D_1^\perp + D_2^\perp)^\perp \subset D_1 \cap D_2.$$

and

$$(D_1 \cap D_2)^\perp \subset ((D_1^\perp + D_2^\perp)^\perp)^\perp.$$

Since $D_1 \cap D_2$ has constant dimension, so does $(D_1^\perp + D_2^\perp)^\perp$, and we must have that $((D_1^\perp + D_2^\perp)^\perp)^\perp = (D_1^\perp + D_2^\perp)$. This implies that

$$(D_1 \cap D_2)^\perp \subset D_1^\perp + D_2^\perp.$$

Since we always have that

$$D_1^\perp + D_2^\perp \subset (D_2 \cap D_1)^\perp,$$

we can conclude that, restricted to U,

$$(D_1 \cap D_2)^\perp = D_1^\perp + D_2^\perp.$$

$\square$.

Example: The first example illustrates that D can be a proper subset of $(D^\perp)^\perp$. Let (x,y) be the standard coordinates on $\mathcal{R}^2$. Consider the distribution defined by $D := \text{span}\{y^2\frac{\partial}{\partial x}\}$. The perp of this distribution is defined by

$$D^\perp := \{a(x,y)dx + b(x,y)dy \in \Omega^1(\mathcal{R}^2) \mid v \lrcorner (a(x,y)dx + b(x,y)dy)\ \forall v \in D\}.$$

Consequently, if $a(x,y)dx + b(x,y)dy \in D^\perp$, then for every $c(x,y) \in C^\infty(\mathcal{R}^2)$, the equation

$$(c(x,y)y^2\frac{\partial}{\partial x}) \lrcorner (a(x,y)dx + b(x,y)dy) \equiv 0$$

must hold. This can be rewritten as

$$a(x,y)c(x,y)y^2 + b(x,y) \cdot 0 \equiv 0.$$

The coefficients $a(x,y) \equiv 0$, $b(x,y) \equiv 1$ obviously satisfy this equation, so $dy \in D^\perp$. Furthermore, any $\omega \in D^\perp$ with $a(x,y) \equiv 0$ can be written as a scalar multiple of dy. We still need to determine if there are any elements of $D^\perp$ with nonzero $a(x,y)$ coefficients. If so, then the function $a(x,y)$ must satisfy $a(x,y)c(x,y)y^2 \equiv 0$ for all $c(x,y) \in C^\infty(M)$. In particular, this equation must hold when $c(x,y) \equiv 1$. Wherever $y \neq 0$, the function $a(x,y) = 0$, since this holds on a dense subset of $\mathcal{R}^2$, the continuity of $a(x,y)$ implies that the zero function $a(x,y) \equiv 0$ is the only solution. So $D^\perp = \text{span}\{dy\}$. A similar argument shows that $(D^\perp)^\perp = \text{span}\{\frac{\partial}{\partial x}\}$. The distribution D is a proper subspace of $(D^\perp)^\perp$ because dy cannot be written as any $c(x,y)(y^2dx)$. ◇

Example: The second example illustrates that $D_1^\perp + D_2^\perp$ can be a proper subset of $(D_2 \cap D_1)^\perp$. Let (x,y,z) be the standard coordinates on $\mathcal{R}^3$. Let $D_1 := \text{span}\{\frac{\partial}{\partial y}, x\frac{\partial}{\partial x} + \frac{\partial}{\partial z}\}$ and $D_2 := \text{span}\{\frac{\partial}{\partial y}, \frac{\partial}{\partial z}\}$. Then $D_1^\perp = \text{span}\{dx - xdz\}$, $D_2^\perp = \text{span}\{dx\}$, and $D_1 \cap D_2 = \text{span}\{\frac{\partial}{\partial y}\}$. Computing $(D_1 \cap D_2)^\perp$ and $D_1^\perp + D_2^\perp$, we find that $(D_1 \cap D_2)^\perp = \text{span}\{dx, dz\}$ while $D_1^\perp + D_2^\perp = \text{span}\{dx - xdz, dx\} = \text{span}\{dx, xdz\}$. The distribution $\text{span}\{dx, xdz\}$ is a proper submodule of $\text{span}\{dx, dz\}$ since dz is not an element of $\text{span}\{dx, xdz\}$. ◇

7.2.5 Closure Properties of Distributions and Codistributions

We can define operations on smooth distributions and codistributions which are induced by the Lie bracket and Lie derivative. In this section, we will define these induced operations, and examine some closure properties which are related to them.

Let Δ and Γ be two smooth, nonzero distributions and I a smooth codistribution. We will define the submodules $[\Delta, \Gamma]$ and $L_\Delta I$ as follows

1. $[\Delta, \Gamma] := \text{span}\{v \in TM \mid v = [r,s] \text{ for some } r \in \Delta, s \in \Gamma\}$.
2. $L_\Delta I := \text{span}\{\omega \in T^*M \mid \omega = L_r\alpha \text{ for some } r \in \Delta, \alpha \in I\}$.

Lemma 7.2.5 *Let $v_1, \ldots, v_n$ be a collection of linearly independent vector fields which span a distribution Δ and $r_1, \ldots, r_m$ a collection of linearly independent vector fields which span a distribution Γ. Then*

$$[\Delta, \Gamma] \subset span\{[v_i, r_j], v_k, r_l\} \quad i,k = 1, \ldots, n \quad j,l = 1, \ldots, m.$$

Furthermore, if Δ and Γ have constant dimension in a neighborhood U of a point p, then there exists a neighborhood $V \subset U$ of p on which

$$[\Delta, \Gamma] = span\{[v_i, r_j], v_k, r_l\} \quad i,k = 1, \ldots, n, \quad j,l = 1, \ldots, m.$$

Proof: The proof of the first statement follows from the following identity. For any $v, r \in TM$ and any $a, b \in C^\infty(M)$,

$$[av, br] = ab[v, r] + (L_v b)ar - (L_r a)bv. \tag{7.15}$$

Let $\hat{\Delta} :=$ span$\{[v_i, r_j], v_k, r_l\}$. For any $v \in \Delta$, $r \in \Gamma$, and $[v, r] = [a^i v_i, b^j r_j]$, by equation 7.15 we have that

$$[a^i v_i, b^j r_j] = a^i b^j [v_i, r_j] + (L_{v_i} b^j) a^i r_j - (L_{r_j} a^i) b^j v_i \in \hat{\Delta}.$$

To prove the second statement, we need to show that locally, $[\Delta, \Gamma] \subset \hat{\Delta}$. Pick any pair of basis vectors $v_i \in \Delta, r_j \in \Gamma$. By definition, $[v_i, r_j] \in [\Delta, \Gamma]$. Now pick any $a \in C^\infty(M)$ such that locally $a \neq 0$ and $L_{r_j} a \neq 0$. The vector field $[av_i, r_j] \in [\Delta, \Gamma]$ again by definition. Consequently, the linear combination

$$\frac{a}{L_r a}[v_i, r_j] - \frac{1}{L_r a}[av_i, r_j] = r_j \in [\Delta, \Gamma].$$

The same argument can be used to show that locally $v_i \in [\Delta, \Gamma]$. □

Lemma 7.2.6 *Let $v_1, \ldots, v_n$ be a collection of linearly independent vector fields which span a distribution Δ and $\omega^1, \ldots, \omega^m$ a collection of linearly independent covector fields which span a codistribution I. If $\Delta \subset I^\perp$, then*

$$L_\Delta I \subset span\{L_{v_i}\omega^j, \omega^k\} \;\; i, k = 1, \ldots, n \;\; j = 1, \ldots, m.$$

Furthermore, if the dimension of Δ is greater than zero in a neighborhood U of a point p, then there exists a neighborhood $V \subset U$ of p on which

$$L_\Delta I = span\{L_{v_i}\omega^j, \omega^k\} \;\; i, k = 1, \ldots, n \;\; j = 1, \ldots, m.$$

Proof: The proof of the first statement uses the following identity. For any $v \in V(M)$, any $\omega \in \Omega^1(M)$, and any $a, b \in C^\infty(M)$,

$$L_{av}(b\omega) = abL_v\omega + (L_v b)a\omega + (v \lrcorner \omega)bda. \tag{7.16}$$

Let $\hat{I} :=$ span$\{L_{v_i}\omega^j, \omega^k\}$, $\;i, k = 1, \ldots, n$, and $j = 1, \ldots, m$. For any $\omega \in I$ and any $v \in \Delta$, $L_v\omega = L_{a^i v_i} b_j \omega^j$ and by equation 7.16 we have that

$$L_{a^i v_i} b_j \omega^j = a^i b^j L_{v_i}\omega^j + (L_{v_i} b^j) a^i \omega^j + b^j (v_i \lrcorner \omega^j) da^i \in \hat{I}$$

which is in $\hat{I}$ since by assumption $v_i \lrcorner \omega^j \equiv 0$.

To prove the second statement, we have to show that locally $\hat{I} \subset L_\Delta I$. Pick any nonzero $v_i \in \Delta$ and any $\omega^j \in I$. By definition, $L_{v_i}\omega^j \in L_\Delta I$. Now pick any $b \in C^\infty(M)$ such that locally $b \neq 0$ and $L_{v_i} b \neq 0$. Then the linear combination

$$\frac{b}{L_r b} L_{v_i}\omega^j - \frac{1}{L_{v_i} b} L_{v_i}(b\omega^j) = \omega^j \in L_\Delta I.$$

□

Lemma 7.2.7 *Suppose that J is a codistribution which is a subspace of $\Delta^\perp \cap I$, then the following statements are equivalent*

1. $L_\Delta J \subset I$
2. $L_{I^\perp} J \subset \Delta^\perp$
3. $[\Delta, I^\perp] \subset J^\perp$

Proof: (1 $\Leftrightarrow$ 2) Since $J \subset \Delta^\perp \cap I$, $\Delta \lrcorner J \equiv 0$ and $I^\perp \lrcorner J \equiv 0$. Therefore, $L_\Delta J = \Delta \lrcorner dJ$ and $L_{I^\perp} J = I^\perp \lrcorner dJ$.

If $L_\Delta J = \Delta \lrcorner dJ \subset I$, then we must have

$$I^\perp \lrcorner (\Delta \lrcorner dJ) = \Delta \lrcorner (I^\perp \lrcorner dJ) \equiv 0$$

which implies that $I^\perp \lrcorner dJ \subset \Delta^\perp$. A symmetric argument gives the reverse implication.

(2 $\Leftrightarrow$ 3) Using Cartan's formula and the facts that $J \subset I$, $J \subset \Delta^\perp$,

$$\Delta \lrcorner I^\perp \lrcorner dJ = [\Delta, I^\perp] \lrcorner J \equiv 0$$

which implies that

$$[\Delta, I^\perp] \subset J^\perp.$$

□

Lemma 7.2.8 *Given a smooth codistribution I and a smooth distribution Δ, there exists a unique maximal codistribution $J^* \subset I \cap \Delta^\perp$ which satisfies $L_\Delta J^* \subset I$. Furthermore, $J^* = [\Delta, I^\perp]^\perp \cap I \cap \Delta^\perp$.*

Proof: Consider the collection $\mathcal{K}$ of all codistributions which satisfy this property. The collection contains $\{0\}$, so it is not empty. Let $J_1, J_2 \in \mathcal{K}$, then

$$L_\Delta J_1 + L_\Delta J_2 = L_\Delta (J_1 + J_2) \subset I,$$

so $\mathcal{K}$ is closed under subspace addition. Therefore, it must contain a unique element of maximal dimension.

The lemma assumes that $J^* \subset I \cap \Delta^\perp$, and using Lemma 7.2.7, we know that $[\Delta, I^\perp] \subset J^{*\perp}$, or equivalently, that $J^* \subset [\Delta, I^\perp]^\perp$. Therefore, we must have $J^* \subset [\Delta, I^\perp]^\perp \cap I \cap \Delta^\perp$. Let $\hat{J} = [\Delta, I^\perp]^\perp \cap I \cap \Delta^\perp$. Using Cartan's formula, we have that

$$[\Delta, I^\perp] \lrcorner \hat{J} = \Delta \lrcorner I^\perp \lrcorner d(\hat{J}) \equiv 0.$$

From this, we infer that

$$L_\Delta(\hat{J}) \subset I.$$

From the maximality of J^*, we conclude that $\hat{J} \subset J^*$. Consequently, we must have that

$$[\Delta, I^\perp]^\perp \cap I \cap \Delta^\perp = J^*.$$

□

Lemma 7.2.9 *Given two smooth codistributions I and J satisfying $J \subset I$, there exists a unique maximal distribution $\Delta^* \subset I^\perp$ which satisfies*

$$L^*_\Delta J \subset I. \tag{7.17}$$

Furthermore, $(I + L_{I^\perp} J)^\perp = \Delta^*$.

Proof: Let $\mathcal{S}$ denote the set of all distributions which satisfy equation 7.17. This set is not empty since $\{0\}$ always satisfies the condition. Suppose that Δ_1 and Δ_2 are both in $\mathcal{S}$. It follows from this that

$$L_{\Delta_1} J + L_{\Delta_2} J = (\Delta_1 + \Delta_2) \lrcorner\, dJ = L_{(\Delta_1+\Delta_2)} J \subset I,$$

so the collection $\mathcal{S}$ is closed under subspace addition. Therefore, it contains a unique maximal element.

Let Δ^* denote the maximal element of $\mathcal{S}$. Using Lemma 7.2.7, we know that

$$L_{\Delta^*} J \subset I \Leftrightarrow L_{I^\perp} J \subset \Delta^{*\perp}.$$

Since the lemma assumes that $I \subset \Delta^{*\perp}$, we must have

$$(L_{I^\perp} J + I) \subset \Delta^{*\perp} \rightarrow \Delta^* \subset (L_{I^\perp} J + I)^\perp.$$

Define $K := L_{I^\perp} J + I$. Clearly, $L_{I^\perp} J \subset K$. Using Lemma 7.2.7, this implies that $L_{K^\perp} J \subset I$. From the maximality of Δ^*, we must have $K^\perp \subset \Delta^*$. From this, we conclude that $(I + L_{I^\perp} J)^\perp = \Delta^*$.
□

Involutivity

A distribution is said to be <u>involutive</u> if and only if it is closed under the Lie bracket operation, so that $[\Delta, \Delta] \subset \Delta$. A codistribution I is <u>involutive</u> if and only if $L_{I^\perp} I \subset I$.

Lemma 7.2.10 *The following facts concerning involutivity hold*

1. *Given two involutive codistributions I and J, the codistribution $I + J$ is also involutive.*
2. *Given two involutive distributions Δ and Γ, the distribution $\Delta \cap \Gamma$ is also involutive.*
3. *Given a codistribution I, there exists a unique maximal involutive codistribution $\hat{I}$ which is contained in I.*
4. *Given a distribution Δ, there exists a unique minimal involutive distribution $\hat{\Delta}$ which contains Δ.*

Proof: These results follow directly from the definition of involutivity. □

A distribution Δ is said to be <u>completely integrable</u> at a point $p \in M$ if and only if it has constant dimension on a neighborhood $U \subset M$ of p and at each point $q \in U$, there exists a submanifold N of U which contains q whose tangent space $T_r N = \Delta(r)$ at each point $r \in N$. The submanifold N is called an <u>integral manifold</u> of the distribution Δ.

Similarly, a codistribution Ω is said to be completely integrable at a point p if and only if it has constant dimension on a neighborhood $U \subset M$ of p and at each point $q \in U$, there exists a submanifold N of U which contains q whose cotangent space $T_r^* N = \Omega(r)$ at each point $r \in N$. The submanifold N is called an integral manifold of the codistribution Ω.

The following fundamental result, called the Frobenius theorem, provides us with a condition under which a distribution or codistribution is completely integrable.

Theorem 7.2.3 (Frobenius Theorem) *An m-dimensional distribution Δ is completely integrable on a neighborhood $U \subset M$ if and only if it is involutive on U. Furthermore, if Δ is completely integrable, then there exists a set of local coordinates $x^1, \ldots, x^m, x^{m+1}, \ldots x^n$ defined on U such that*

$$\Delta = span\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^m}\}.$$

An m-dimensional codistribution Ω is completely integrable on a neighborhood $U \subset M$ if and only if it is involutive on U. Furthermore, if Ω is completely integrable, then there exists a set of local coordinates $x^1, \ldots, x^m, x^{m+1}, \ldots x^n$ defined on U such that

$$\Omega = span\{dx^1, \ldots, dx^m\}.$$

Proof: See Spivak [6]. □

Chapter 8

Exterior Differential Systems

In this chapter, we will extend the exterior algebra presented in Chapter 6 to a smooth manifold. We will then discuss exterior differential systems which are the counterparts in this setting to systems of exterior equations over a vector space.

8.1 The Exterior Algebra on a Manifold

The space of all forms on a manifold M,

$$\Omega(M) = \Omega^0(M) \oplus \ldots \oplus \Omega^n(M),$$

together with the wedge product is called the exterior algebra on M. An ideal of this algebra is defined in Section 6.1 as a subspace $I \subset \Omega(M)$ which satisfies the requirement that if $\alpha \in I$, then $\alpha \wedge \beta \in I$ for any $\beta \in \Omega(M)$. We will call this the algebraic ideal generated by Σ.

We are also interested in what happens when we perform exterior differentiation on the elements of the ideal.

Definition 8.1.1 *An ideal $I \subset \Omega(M)$ is said to be closed with respect to exterior differentiation if and only if*

$$\alpha \in I \Longrightarrow d\alpha \in I$$

or more compactly $dI \subset I$. An ideal which is closed with respect to exterior differentiation is called a differential ideal.

A finite collection of forms $\Sigma := \{\alpha^1, \ldots, \alpha^K\}$ generates an algebraic ideal

$$I_\Sigma := \{\omega \in \Omega(M) \mid \omega = \sum_{i=1}^{K} \theta^i \wedge \alpha^i \text{ for some } \theta^i \in \Omega(M)\}.$$

We can also talk about the differential ideal generated by Σ.

Definition 8.1.2 *Let S_d denote the collection of all differential ideals containing Σ. The differential ideal generated by Σ is defined as*

$$\mathcal{I}_\Sigma := \bigcap_{I \in S_d} I.$$

Theorem 8.1.1 *Let Σ be a finite collection of forms, and let $\mathcal{I}_\Sigma$ denote the differential ideal generated by Σ. Define the collection*

$$\Sigma' = \Sigma \cup d\Sigma$$

and denote the algebraic ideal which it generates by $I_{\Sigma'}$. Then

$$\mathcal{I}_\Sigma = I_{\Sigma'}.$$

Proof: By definition, $\mathcal{I}_\Sigma$ is closed with respect to exterior differentiation, so $\Sigma' \subset \mathcal{I}_\Sigma$. Consequently, $I_{\Sigma'} \subset \mathcal{I}_\Sigma$. Conversely, the ideal $I_{\Sigma'}$ is closed with respect to exterior differentiation and contains Σ by construction. Therefore, from the definition of $\mathcal{I}_\Sigma$, we have that $\mathcal{I}_\Sigma \subset I_{\Sigma'}$. □

8.2 Exterior Differential Systems

In Section 6.2, we introduced systems of exterior equations on a vector space V and characterized their solutions as subspaces of V. We are now ready to define a similar notion for a collection of differential forms defined on a manifold M. The basic problem will be to study the integral submanifolds of M which satisfy the constraints represented by the exterior differential system.

Definition 8.2.1 *An exterior differential system is a finite collection of equations*

$$\alpha^1 = 0, \ldots, \alpha^r = 0, d\alpha^1 = 0, \ldots, d\alpha^r = 0$$

where each $\alpha^i \in \Omega^k(M)$ is a smooth k-form. A solution to an exterior differential system is any submanifold N of M which satisfies $\alpha^i(p)|_{T_pN} \equiv 0$ and $d\alpha^i(p)|_{T_pN} \equiv 0$ for all $p \in N$ and all $i \in 1, \ldots, r$. An integral k-plane is a point $\Delta_p \in G^n_k(M)$ which satisfies the equations $\alpha^i(p)|_{\Delta_p} \equiv 0$, $d\alpha^i(p)|_{\Delta_p} \equiv 0$ for all $i \in 1, \ldots, r$.

Theorem 8.2.1 *Given an exterior differential system*

$$\alpha^1 = 0, \ldots, \alpha^K = 0 \tag{8.1}$$

and the corresponding differential ideal $\mathcal{I}_\Sigma$ generated by the collection of forms

$$\Sigma := \{\alpha^1, \ldots, \alpha^K, d\alpha^1, \ldots, d\alpha^K\}, \tag{8.2}$$

an integral submanifold N of M solves the system of exterior equations if and only if it also solves the equation $\pi = 0$ for every $\pi \in \mathcal{I}_A$.

Proof: If an integral submanifold N of M is a solution to Σ, then for all $x \in N$ and all $i \in 1, \ldots, K$,

$$\alpha^i(x)|_{T_xN} \equiv 0.$$

Taking the exterior derivative gives

$$d\alpha^i(x)|_{T_xN} \equiv 0.$$

Therefore, the submanifold also satisfies the exterior differential system

$$\alpha^1 = 0, \ldots, \alpha^K = 0, d\alpha^1 = 0, \ldots, d\alpha^K = 0.$$

From Theorem 8.1.1, we know that the differential ideal generated by Σ is equal to the algebraic ideal generated by the above system. Therefore, from Theorem 6.2.1, we know that N will also be a solution for every element of $\mathcal{I}_\Sigma$.

Conversely, if N solves the equation $\pi = 0$ for every $\pi \in \mathcal{I}_\Sigma$, then in particular it must solve Σ. □

The above theorem allows us to either work with the generators of an ideal or with the ideal itself. In fact, some authors define exterior differential systems as differential ideals of $\Omega(M)$.

Because a set of generators Σ generates both a differential ideal $\mathcal{I}_\Sigma$ and an algebraic ideal I_Σ, we can define two different notions of equivalence for exterior differential systems.

Definition 8.2.2 *Two exterior differential systems, Σ_1 and Σ_2, are said to be algebraically equivalent if and only if they generate the same algebraic ideal, i.e., $I_{\Sigma_1} = I_{\Sigma_2}$.*

Definition 8.2.3 *Two exterior differential systems, Σ_1 and Σ_2, are said to be equivalent if and only if they generate the same differential ideal, i.e., $\mathcal{I}_{\Sigma_1} = \mathcal{I}_{\Sigma_2}$.*

Intuitively, we want to think of two exterior differential systems as equivalent if they have the same solution set. Therefore, we will usually discuss equivalence in terms of this second definition.

8.3 The Cauchy Characteristic Distribution

In Chapter 6, we introduced the associated and retracting spaces for a system of exterior equations. We will now introduce analogous concepts for an exterior differential system.

The associated space and retracting space of an ideal in $\Omega(M)$ are defined pointwise as in Section 6.2. The associated space of a differential ideal $\mathcal{I}$ is called the Cauchy characteristic distribution and is denoted by $A(\mathcal{I})$. The main result of this section is a theorem which uses the Cauchy characteristic to find a set of generators Σ' which is equivalent to Σ and can be described using a minimal set of coordinate functions. Before stating this theorem, we will first prove that the Cauchy characteristic distribution is involutive.

Theorem 8.3.1 *The Cauchy characteristic distribution is involutive.*

Proof: Let $\mathcal{I}$ be a differential ideal. By definition, a vector field $v \in V(M)$ is contained in the Cauchy characteristic distribution $A(\mathcal{I})$ if and only if

$$v \lrcorner \mathcal{I} \subset \mathcal{I}.$$

Given any two vector fields $v, r \in A(\mathcal{I})$ and any k-form $\omega \in \mathcal{I}$, we can apply Lemma 7.9 to compute

$$[v,r] \lrcorner \omega = L_v(r \lrcorner \omega) - r \lrcorner L_v \omega$$

$$= v \lrcorner d(r \lrcorner \omega) + d(v \lrcorner r \lrcorner \omega) - r \lrcorner v \lrcorner d\omega - r \lrcorner d(v \lrcorner \omega).$$

Since $\mathcal{I}$ is closed with respect to the d operator and the $v \lrcorner$ and $r \lrcorner$ operators, we find that

$$[v,r] \lrcorner \omega \in \mathcal{I}.$$

Since ω was arbitrary, this implies that

$$[v,r] \lrcorner \mathcal{I} \subset \mathcal{I}.$$

Hence, for any $v, r \in A(\mathcal{I})$, we have that $[v,r] \in A(\mathcal{I})$, so $A(\mathcal{I})$ is involutive. □

If the Cauchy characteristic distribution has constant dimension s on a neighborhood U, then the Frobenius theorem says that there exists a set of $n-s$ smooth functions $y^1, \ldots, y^{n-s}$ whose differentials span the retracting space $C(\mathcal{I}) = A(\mathcal{I})^\perp$ on this neighborhood. If we add an additional set of functions $x^1, \ldots, x^s$ so that the y's and x's taken together form a coordinate chart over U, then the retraction theorem, Theorem 6.2.2, says that at each point $p \in U$, we can construct a set of generators for $\mathcal{I}$ which only involves the 1-forms $\{dy^1, \ldots, dy^{n-s}\}$. Each k-form generator in this set can be written with respect to the dy as

$$\omega = \sum_{I \in K} a_I(x,y) dy^I. \tag{8.3}$$

The following theorem considerably strengthens this result by showing that there exists a set of generators of the form 8.3 in which the coefficients are also only functions of the y coordinates.

Theorem 8.3.2 *Let $\mathcal{I}$ be a finitely generated differential ideal whose retracting space $C(\mathcal{I})$ has constant dimension $s = n - p$. Then there is a neighborhood in which there are coordinates $(x^1, \ldots, x^p, y^1, \ldots, y^s)$ such that $\mathcal{I}$ has a set of generators that are forms in $y^1, \ldots, y^s$ and their differentials.*

Before proving this theorem, it may prove helpful to discuss its geometric significance. Suppose that $\mathcal{I}_{\mathcal{M}}$ is an exterior differential system which is defined over a smooth manifold M. If the Cauchy characteristic distribution has constant dimension in a neighborhood U of some point $p \in M$, then there exists a foliation of smooth manifolds in U and a coordinate chart defined on a neighborhood $V \subset U$ of p with component functions $(x^1, \ldots, x^p, y^1, \ldots, y^s)$. Each leaf of the foliation is described by the equations $y^1 = c^1, \ldots, y^s = c^s$ for some vector of constants $(c^1, \ldots, c^s)$. Restricted to the neighborhood V, this relationship defines a smooth surjection $\gamma : V \to \mathcal{R}^s$. If $\omega = \sum_{I \in \mathcal{A}} a_I(y) dy^I$ is any form on $\mathcal{R}^s$, then its pullback to V will still be described in local coordinates by $\gamma^* \omega = \sum_{I \in \mathcal{A}} a_I(y) dy^I$. Therefore, Theorem 8.3.2 can be interpreted as saying that there exists a set of generators for $\mathcal{I}$ which locally coincide with a set of forms which have been pulled back from $\mathcal{R}^s$.

The proof of Theorem 8.3.2 will make use of the following lemma.

Lemma 8.3.1 *Let $s, n,$ and p be positive integers satisfying the equation $s = n - p$. Let $U \subset \mathcal{R}^n$ be an open subset with local coordinates $x^1, \ldots, x^p$, $y^1, \ldots, y^s$, and let $B_1, \ldots, B_p$ be a set of p smooth functions taking U to $\mathcal{R}^{r \times r}$. There exists a function $D : U \to \mathcal{R}^{r \times r}$ which satisfies the partial differential equation*

$$\frac{\partial D^i_k}{\partial x^j} = \sum_{l=1}^{r} D^i_l B^l_{jk} \tag{8.4}$$

if and only if

$$\frac{\partial B^l_{tk}}{\partial x^j} - \frac{\partial B^l_{jk}}{\partial x^t} + \sum_{v=1}^{r} (B^l_{jv} B^v_{tk} - B^l_{tv} B^v_{jk}) = 0 \tag{8.5}$$

for $1 \leq l, k \leq r$ and $1 \leq j, t \leq p$.

Proof: The necessity of this condition can be seen taking the second partial derivatives of equation 8.4 and then using equation again 8.4 to substitute for $\frac{\partial D^i_l}{\partial x^t}$.

$$\frac{\partial^2 D^i_k}{\partial x^t \partial x^j} = \sum_{l=1}^{r} \frac{\partial D^i_l}{\partial x^t} B^l_{jk} + \sum_{l=1}^{r} D^i_l \frac{\partial B^l_{jk}}{\partial x^t}$$

$$= \sum_{l=1}^{r} \sum_{v=1}^{r} D^i_v B^v_{tl} B^l_{jk} + \sum_{l=1}^{r} D^i_l \frac{\partial B^l_{jk}}{\partial x^t}.$$

Permuting the t and j indices and subtracting gives

$$\frac{\partial^2 D^i_k}{\partial x^j \partial x^t} - \frac{\partial^2 D^i_k}{\partial x^t \partial x^j} = \sum_{l=1}^{r} \left(D^i_l \left(\sum_{v=1}^{r} B^l_{jv} B^v_{tk} + \frac{\partial B^l_{tk}}{\partial x^j} \right) - D^i_l \left(\sum_{v=1}^{r} B^l_{tv} B^v_{jk} + \frac{\partial B^l_{jk}}{\partial x^t} \right) \right) = 0$$

Since the matrix D^i_l is nonsingular, this implies equation 8.5.

To prove the sufficiency of this condition, we will work on the space $FU := U \times R^{r \times r}$. At any point (x, y, D), the tangent space $T_{(x,y,D)} FU$ is locally spanned by the vector fields

$$T_{(x,y,D)} FU = \text{span} \left\{ \frac{\partial}{\partial D^1_1}, \ldots, \frac{\partial}{\partial D^1_r}, \frac{\partial}{\partial D^2_1}, \ldots, \frac{\partial}{\partial D^r_r}, \frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^s}, \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^p} \right\}.$$

We will consider the distribution Δ defined with respect to this basis by the last p columns of the matrix

$$\begin{bmatrix} I_{r^2 \times r^2} & 0 & \sum_{v=1}^{r} D^i_v B^v_{1k} & \cdots & \sum_{v=1}^{r} D^i_v B^v_{pk} \\ 0 & I_{s \times s} & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & & 1 \end{bmatrix}.$$

The inverse of this matrix is given by

$$\begin{bmatrix} I_{r^2\times r^2} & 0 & -\sum_{v=1}^{r} D_v^i B_{1k}^v & \cdots & -\sum_{v=1}^{r} D_v^i B_{pk}^v \\ 0 & I_{s\times s} & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & & 1 \end{bmatrix}$$

which implies that $\Delta^\perp$ is spanned by the covectors

$$\Delta^\perp = \text{span}\left\{ dy^1, \ldots, dy^s, dD_1^1 - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 B_{j1}^v dx^j, \ldots, dD_r^r - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^r B_{jr}^v dx^j \right\}. \tag{8.6}$$

Taking the exterior derivative of the $dD_1^1 - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 B_{j1}^v dx^j \in \Delta^\perp$, we find that

$$d(dD_1^1 - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 B_{j1}^v dx^j) = -\sum_{j=1}^{p}\sum_{v=1}^{r} B_{j1}^v dD_v^1 \wedge dx^j - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 dB_{j1}^v \wedge dx^j. \tag{8.7}$$

Examining the first term on the right, we find that

$$-\sum_{j=1}^{p}\sum_{v=1}^{r} B_{j1}^v dD_v^1 \wedge dx^j \quad = \quad -\sum_{j=1}^{p}\sum_{v=1}^{r} B_{j1}^v \left(\sum_{l=1}^{p}\sum_{q=1}^{r} D_q^1 B_{lv}^q\right) dx^l \wedge dx^j \text{ mod } \Delta^\perp.$$

Similarly, expanding the second term on the right gives

$$-\sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 dB_{j1}^v \wedge dx^j \quad = \quad -\sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 \left(\sum_{l=1}^{p} \frac{\partial B_{j1}^v}{\partial x^l}\right) dx^l \wedge dx^j \text{ mod } \Delta^\perp.$$

Collecting terms, we find that

$$d(dD_1^1 - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 B_{j1}^v dx^j) = -\sum_{j=1}^{p}\sum_{l=1}^{p}\left(\sum_{v=1}^{r} D_v^1 \left(\frac{\partial B_{j1}^v}{\partial x^l} + \sum_{q=1}^{r} B_{j1}^q B_{lq}^v\right)\right) dx^l \wedge dx^j \text{ mod } \Delta^\perp$$

$$= \sum_{1\le l<j\le p} \left(\sum_{v=1}^{r} D_v^1 \left(\frac{\partial B_{j1}^v}{\partial x^l} - \frac{\partial B_{l1}^v}{\partial x^j} + \sum_{q=1}^{r} B_{j1}^q B_{lq}^v - \sum_{q=1}^{r} B_{l1}^q B_{jq}^v\right)\right) dx^l \wedge dx^j \text{ mod } \Delta^\perp.$$

If equation 8.5 is satisfied, then we find that

$$d(dD_1^1 - \sum_{j=1}^{p}\sum_{v=1}^{r} D_v^1 B_{j1}^v dx^j) = 0 \text{ mod } \Delta^\perp.$$

If we carry out these computations for each basis element of $\Delta^\perp$, we find that the codistribution satisfies the Frobenius condition. Consequently, the Frobenius theorem guarantees that there exists a set of $(r^2 + s)$ smooth functions whose differentials span $\Delta^\perp$. Furthermore, since we know that $\{dy^1, \ldots, dy^s\} \subset \Delta^\perp$, we can select a set of functions $\hat{D}_j^i$ with $1 \le i, j \le r$ such that $\Delta^\perp = \text{span}\{dy^1, \ldots, dy^s, d\hat{D}_1^1, \ldots d\hat{D}_r^r\}$. Finally, we know that none of the covectors $\{dx^1, \ldots, dx^p\}$ lie in the span of $\Delta^\perp$, so

$$T^*M = \text{span}\{dx^1, \ldots, dx^p, dy^1, \ldots, dy^s, d\hat{D}_1^1, \ldots, d\hat{D}_r^r\}.$$

The coordinate transformation taking the coordinates $(D,y,x) \to (\hat{D},y,x)$ is of the form

$$\begin{aligned}(D,y,x) &\to \hat{D}(x,y,D)\\(D,y,x) &\to y\\(D,y,x) &\to x.\end{aligned}$$

Each integral manifold of Δ can be described by the equations

$$\begin{aligned}\hat{D}(D,y,x) &= c_1\\ y &= c_2\end{aligned}$$

where $c_1 \in \mathcal{R}^{r\times r}$ and $c_2 \in \mathcal{R}^s$ are constants. Taking the differentials of the components of $\hat{D}$, we obtain

$$d\hat{D}_t^w = \sum_{i,k=1}^{r} \frac{\partial \hat{D}_t^w}{\partial D_k^i} dD_k^i + \sum_{q=1}^{s} \frac{\partial \hat{D}_t^w}{\partial y^q} dy^q + \sum_{l=1}^{p} \frac{\partial \hat{D}_t^w}{\partial x^l} dx^l \in \Delta^\perp.$$

The matrix $\frac{\partial \hat{D}}{\partial D}$ is always nonsingular, so it can be inverted to obtain the equations

$$\sum_{w,t=1}^{r} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} d\hat{D}_t^w = dD_k^i + \sum_{w,t=1}^{r} \left(\sum_{q=1}^{s} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} \frac{\partial \hat{D}_t^w}{\partial y^q} dy^q + \sum_{l=1}^{p} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} \frac{\partial \hat{D}_t^w}{\partial x^l} dx^l \right) \in \Delta^\perp.$$

Comparing this expression to equation 8.6, we find that

$$\sum_{w,t=1}^{r} \sum_{l=1}^{p} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} \frac{\partial \hat{D}_t^w}{\partial x^l} dx^l = -\sum_{l=1}^{p} \sum_{v=1}^{r} D_v^i B_{lk}^v dx^l. \tag{8.8}$$

Furthermore, because $\frac{\partial \hat{D}}{\partial D}$ is nonsingular, the implicit function theorem ensures that locally there exists a function $f : U \to \mathcal{R}^{r\times r}$ satisfying $\hat{D}(f(y,x),y,x) = c_1$. Taking the exterior derivative of this equation, we find that

$$\sum_{l=1}^{p} \frac{\partial \hat{D}_t^w}{\partial x^l} dx^l + \sum_{q=1}^{s} \frac{\partial \hat{D}_t^w}{\partial y^q} dy^q + \sum_{i,k=1}^{r} \left(\frac{\partial \hat{D}_k^w}{\partial D_k^i}\right) \left(\frac{\partial f_k^i}{\partial x^l} dx^l + \frac{\partial f_k^i}{\partial y^q} dy^q\right) = 0$$

which implies that

$$\sum_{l=1}^{p} \sum_{w,t=1}^{r} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} \frac{\partial \hat{D}_k^i}{\partial x^l} dx^l + \sum_{q=1}^{s} \sum_{w,t=1}^{r} \left(\frac{\partial \hat{D}}{\partial D}\right)^{-1}_{\substack{ti\\wk}} \frac{\partial \hat{D}_k^i}{\partial y^q} dy^q + \sum_{l=1}^{p} \frac{\partial f_k^i}{\partial x^l} dx^l + \sum_{q=1}^{s} \frac{\partial f_k^i}{\partial y^q} dy^q = 0.$$

Using equation 8.8 and the fact that the 1-forms dx^l are linearly independent, we can conclude that

$$-\sum_{v=1}^{r} f_v^i(y,x) B_{lk}^v + \frac{\partial f_k^i}{\partial x^l} = 0.$$

Hence, the function f is a solution to the partial differential equation 8.4. □

Proof of Theorem 8.3.2 We begin by noting that the involutivity of $A(\mathcal{I})$ ensures that there exists a local coordinate chart of the form $y^1,\ldots,y^s,x^1,\ldots,x^p$ defined such that

$$C(\mathcal{I}) = \text{span}\{dy^1,\ldots,dy^s\}.$$

Furthermore, the retraction theorem, Theorem 6.2, guarantees that we can always find a set of generators for the differential ideal which is contained in $\Lambda(C(\mathcal{I}))$ so we can assume that each generator $\theta^t \in \mathcal{I}_k$ can be written in the form

$$\theta^t = \sum_{I \in \mathcal{A}} a_I^t(x,y) dy^I$$

where $\mathcal{A}$ is the set of all ascending k-tuples in the set $\{1, \ldots, s\}$ and the sum is taken over all such k-tuples.

For each k in the range $1 \leq k \leq s$, we can select a basis of k-forms for the subspace $\mathcal{I} \cap \Lambda^k(C(\mathcal{I})) \subset \Lambda^k(T^*M)$ which we will denote by

$$\mathcal{I} \cap \Lambda^k(C(\mathcal{I})) = \operatorname{span}\{\phi^1, \ldots, \phi^{r_k}\}$$

with each ϕ^m of the form

$$\phi^m = \sum_{I \in \mathcal{A}} b_I^m(x,y) dy^I$$

for $1 \leq m \leq r_k$. Each of the k-form generators must lie within the span of this set. In fact, since each of these subspaces is contained in $\mathcal{I}$ and contains all the k-form generators, the set of all such bases for all $1 \leq k \leq s$ can be used as a new set of generators for the exterior differential system.

Since $\mathcal{I}$ is closed under exterior differentiation and satisfies $v \lrcorner \mathcal{I} \subset \mathcal{I}$ for every $v \in A(\mathcal{I})$, each k-form ϕ^m will satisfy $v \lrcorner d\phi^m \in \mathcal{I}$. In particular, if we compute $\frac{\partial}{\partial x^l} \lrcorner d\phi^m$ with respect to the local coordinates, we get

$$\begin{aligned} \frac{\partial}{\partial x^l} \lrcorner d\phi^m &= \frac{\partial}{\partial x^l} \lrcorner (\sum_{I \in \mathcal{A}} db_I^m(x,y) \wedge dy^I) \\ &= \frac{\partial}{\partial x^l} \lrcorner \sum_{I \in \mathcal{A}} \sum_{i=1}^{p} \frac{\partial b_I^m}{\partial x^i} dx^i \wedge dy^I \\ &+ \frac{\partial}{\partial x^l} \lrcorner \sum_{I \in \mathcal{A}} \sum_{j=1}^{s} \frac{\partial b_I^m}{\partial y^j} dy^j \wedge dy^I \\ &= \sum_{I \in \mathcal{A}} \frac{\partial b_I^m}{\partial x^l} dy^I. \end{aligned}$$

Since each $\frac{\partial}{\partial x^l} \lrcorner d\phi^m \in \mathcal{I}$, there must exist a set of smooth functions $c_{ls}^m(x,y)$ such that

$$\sum_{I \in \mathcal{A}} \frac{\partial b_I^m}{\partial x^l} dy^I = \sum_{I \in \mathcal{A}} \sum_{s=1}^{r_k} c_{ls}^m b_I^s(x,y) dy^I. \tag{8.9}$$

There exists another set of k-forms $\hat{\phi}^q$, $1 \leq q \leq r_k$, which span $\mathcal{I} \cap \Lambda^k(C(\mathcal{I}))$ and whose coefficients are only functions of the y coordinates if and only if there exists a set of smooth functions $h_m^q(x,y)$ which satisfy the equation

$$\hat{\phi}^q = \sum_{m=1}^{r_k} h_m^q(x,y) \phi^m.$$

Written out in coordinates, this equation takes the form

$$\sum_{I \in \mathcal{A}} \hat{b}_I^q(y) dy^I = \sum_{m=1}^{r_k} h_m^q(x,y) \sum_{I \in \mathcal{A}} b_I^m(x,y) dy^I. \tag{8.10}$$

A set of functions h^q_m which satisfies equation 8.10 exists if and only if

$$\frac{\partial}{\partial x^l} \lrcorner\, d(\sum_{m=1}^{r_k} h^q_m(x,y) \sum_{I \in \mathcal{A}} b^m_I(x,y) dy^I) = 0$$

for all $1 \leq l \leq p$. This equation will be satisfied if and only if for each $I \in \mathcal{A}$,

$$\sum_{m=1}^{r_k} b^m_I \frac{\partial h^q_m}{\partial x^l} + \sum_{m=1}^{r_k} h^q_m \frac{\partial b^m_I}{\partial x^l} = 0.$$

Using equation 8.9, this equation can be rewritten as

$$\sum_{m=1}^{r_k} b^m_I \frac{\partial h^q_m}{\partial x^l} + \sum_{m=1}^{r_k} \sum_{t=1}^{r_k} h^q_m c^m_{lt} b^t_I = 0.$$

Changing the summation indexes in the first term and collecting terms gives

$$\sum_{t=1}^{r_k} (\frac{\partial h^q_t}{\partial x^l} + \sum_{m=1}^{r_k} h^q_m c^m_{lt}) b^t_I = 0.$$

Finally, since each of the row vectors $b^1, \ldots, b^s$ is linearly independent, we must have

$$\frac{\partial h^q_t}{\partial x^l} + \sum_{m=1}^{r_k} h^q_m c^m_{lt} = 0$$

for all $1 \leq l \leq p$ and all $1 \leq t, q \leq r_k$.

Lemma 8.3.1 says that these partial differential equations have a solution if and only if for all $1 \leq l, j \leq p$ and all $1 \leq m, t \leq r_k$,

$$\frac{\partial c^m_{jt}}{\partial x^l} - \frac{\partial c^m_{lt}}{\partial x^j} + \sum_{i=1}^{r_k} (c^m_{ji} c^i_{lt} - c^m_{li} c^i_{jt}) = 0. \tag{8.11}$$

Finally, equation 8.11 always holds. This statement can be proven using the equation 7.9 from Lemma 7.2.3

$$L_{\frac{\partial}{\partial x^i}} (\frac{\partial}{\partial x^j} \lrcorner\, d\phi^m) - \frac{\partial}{\partial x^j} \lrcorner\, L_{\frac{\partial}{\partial x^i}} d\phi^m = [\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] \lrcorner\, d\phi^m = 0.$$

Expanding out this equation, we get

$$\frac{\partial}{\partial x^i} \lrcorner\, d(\frac{\partial}{\partial x^j} \lrcorner\, d\phi^m) + d(\frac{\partial}{\partial x^i} \lrcorner\, \frac{\partial}{\partial x^j} \lrcorner\, d\phi^m) - \frac{\partial}{\partial x^j} \lrcorner\, d(\frac{\partial}{\partial x^i} \lrcorner\, d\phi^m) = 0.$$

The second term in this summation is zero because the form $(\frac{\partial}{\partial x^j} \lrcorner\, d\phi^m)$ only involves dy^I terms. Expanding the first and third terms, we find that

$$\frac{\partial}{\partial x^i} \lrcorner\, d(\frac{\partial}{\partial x^j} \lrcorner\, d\phi^m) = \frac{\partial}{\partial x^i} \lrcorner\, d(\sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} c^m_{jv} b^v_I dy^I)$$

$$= \sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} b^v_I \frac{\partial c^m_{jv}}{\partial x^i} dy^I + \sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} \sum_{l=1}^{r_k} c^m_{jv} c^v_{il} b^l_I dy^I$$

and

$$\frac{\partial}{\partial x^j} \lrcorner\, d(\frac{\partial}{\partial x^i} \lrcorner\, d\phi^m) = \frac{\partial}{\partial x^j} \lrcorner\, d(\sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} c^m_{iv} b^v_I dy^I)$$

$$= \sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} b_I^v \frac{\partial c_{iv}^m}{\partial x^j} dy^I + \sum_{I \in \mathcal{A}} \sum_{v=1}^{r_k} \sum_{l=1}^{r_k} c_{iv}^m c_{jl}^v b_I^l dy^I.$$

Subtracting these equations gives

$$\sum_{I \in \mathcal{A}} \left(\sum_{v=1}^{r_k} b_I^v \frac{\partial c_{jv}^m}{\partial x^i} - \sum_{v=1}^{r_k} b_I^v \frac{\partial c_{iv}^m}{\partial x^j} + \sum_{v=1}^{r_k} \sum_{l=1}^{r_k} c_{jv}^m c_{il}^v b_I^l - \sum_{v=1}^{r_k} \sum_{l=1}^{r_k} c_{iv}^m c_{jl}^v b_I^l \right) dy^I =$$

$$\sum_{I \in \mathcal{A}} \sum_{l=1}^{r_k} \left(\frac{\partial c_{jl}^m}{\partial x^i} - \frac{\partial c_{il}^m}{\partial x^j} + \sum_{v=1}^{r_k} c_{jv}^m c_{il}^v - \sum_{v=1}^{r_k} c_{iv}^m c_{jl}^v \right) b_I^l dy^I =$$

$$\sum_{l=1}^{r_k} \left(\frac{\partial c_{jl}^m}{\partial x^i} - \frac{\partial c_{il}^m}{\partial x^j} + \sum_{v=1}^{r_k} c_{jv}^m c_{il}^v - \sum_{v=1}^{r_k} c_{iv}^m c_{jl}^v \right) \phi^l = 0.$$

Since the k-forms ϕ^l are all linearly independent, each coefficient must be identically zero and the result follows. □

Part III

Applications to Nonlinear Control Theory

Chapter 9

Modeling a Control System Using Grassmann Bundles

The introductory presentation in Chapter 2 asserted that a Grassmann bundle of one-dimensional subspaces is the geometric object which correctly models an affine control system on the state-time manifold. The purpose of this chapter is to give a precise description of this bundle and to discuss how both static and dynamic feedback can be viewed geometrically within this framework.

The chapter is divided into four sections. The first section describes a construction which gives a precise definition of the Grassmann bundle which models an affine control system. The second section looks at the local coordinate descriptions of this bundle and shows how they are related to the standard vector field description of an affine nonlinear control system. This section also discusses affine static state feedback, and illustrates how this class of feedback is naturally incorporated as a part of the structure of the Grassmann bundle. The third section discusses dynamic state feedback. The discussion focuses on some of the different classifications of dynamic state feedback and singles out a particular class which will be used in the following sections. The Grassmann bundle prolongation process is then introduced, and the relationship between a bundle prolongation and dynamic state feedback is illustrated. The fourth section discusses how nonaffine control systems can be modeled as Grassmann bundles. This section concludes with a discussion of the possibility of finding a local diffeomorphism between an affine and a nonaffine control system. Some useful references which are related the material in this chapter are the tract by Huijberts [11], the paper by Sluis [21], and the paper by Sluis and Tilbury [22].

9.1 The Grassmann Bundle Model of a Control System

Suppose we are given a state-space manifold M and a smooth distribution Δ defined over the state-time manifold $M \times \mathcal{R}$. We will require that $dt|_{\Delta} \neq 0$ at every point $(p,t) \in M \times \mathcal{R}$, so that Δ is always transverse to the tangent plane of the state-space submanifold $M \times \{t\}$. Our objective is to construct a Grassmann bundle over $M \times \mathcal{R}$ in such a way that the fibre over each point (p,t) consists of the collection of all one-dimensional subspaces of $T_{(p,t)}(M \times \mathcal{R})$ which are contained in Δ. We will denote this bundle by $G_1^{[0,\Delta]}(M \times \mathcal{R})$. To do this, we will use a slightly modified version of the procedure which was used to construct a Grassmann bundle in Chapter 7.

First, we form the index set Λ which consists of all triples $((x,t),F,U)$ where U is an open set on $M \times \mathcal{R}$, (x,t) is a local coordinate chart taking U to $\mathcal{R}^{n+1}$, and F is a local frame for Δ which can be described relative to the basis

$$\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial t} \right\}$$

at every point $(p,t) \in U$ by

$$f_i = \sum_{j=1}^{m} a_i^j(x,t)\frac{\partial}{\partial x^j} + a^t(x,t)\frac{\partial}{\partial t}.$$

Next, we form the disjoint union

$$\mathcal{U} = \bigcup_{((x,t),F,U)\in\Lambda} \{((x,t),F,U)\} \times U_{((x,t),F,U)} \times G_1^{m+1}$$

where $U_{((x,t),F,U)}$ is a copy of the domain of the coordinate chart (x,t). Each point of $\mathcal{U}$ is a triple $(((x,t),F,U),(p,t),l)$ which satisfies $(p,t) \in U$. If the domains of two coordinate charts $((x,t),F_x,U_x)$ and $((y,t),F_y,U_y)$ have a nonempty intersection $U_x \cap U_y \neq \emptyset$, then since F_y and F_x are both frames which locally span Δ, there must exist a unique smooth function $T : U_x \cap U_y \to \mathcal{R}^{(m+1)\times(m+1)}$ which satisfies the equation

$$F_y T = \frac{\partial y}{\partial x}|_{(p,t)} F_x.$$

From this equation, it is apparent that the definition of the function T depends on both the coordinate systems (x,t) and (y,t) and the local frames F_x and F_y and that it is uniquely defined for each pair of charts whose base sets intersect. We will use the function T to define an equivalence relation on $\mathcal{U}$ such that

$$(((x,t),F_x,U_x),(p,t),l_x) \sim (((y,t),F_y,U_y),(q,\tau),l_y) \tag{9.1}$$

if and only if $(p,t) = (q,\tau)$ and $l_y = \Phi(T(p,t))l_x$. In this equation, $\Phi : GL(\mathcal{R}^{m+1}) \to PGL(\mathcal{R}^{m+1})$ is the group homomorphism discussed in Chapter 5. We will denote the equivalence class of the point $(((x,t),F_x,U_x),(p,t),l_x)$ by $[((x,t),F_x,U_x),(p,t),l_x]$. If we endow the set of all such equivalence classes with the quotient topology, then we have a Grassmann bundle with standard fibre G_1^{m+1} and local trivializations given by the mappings

$$t_{((x,t),F_x,U_x)} : \pi^{-1}(U_{((x,t),F_x,U_x)}) \to U_{((x,t),F_x,U_x)} \times G_1^{m+1}$$

$$[((x,t),F_x,U_x),(p,t),l] \to ((p,t),l).$$

Every Grassmann bundle defined over a smooth base manifold is itself a smooth manifold, so it can be given a smooth differentiable structure. We can construct an atlas for the bundle $G_1^{[0,\Delta]}(M\times\mathcal{R})$ by noting that each local trivialization $U_{((x,t),F_x,U_x)} \times G_1^{m+1}$ can be covered by m charts of the form

$$c_i : V_i \subset U_{((x,t),F_x,U_x)} \times G_1^{m+1} \to \mathcal{R}^{n+1} \times \mathcal{R}^m$$

where the function c_i takes $(p,t) \in U_{((x,t),F_x,U_x)} \to (x(p,t),t)$ and $l \in G_1^{m+1}$ to the ith standard coordinate chart on G_1^{m+1}. Since the bases of the local trivializations cover $M \times \mathcal{R}$, we can form an atlas on the bundle $G_1^{[0,\Delta]}(M)$ by taking the compositions $\left(c_i \circ t_{((x,t),F_x,U_x)}\right)|_{t^{-1}_{((x,t),F_x,U_x)}(V_i)}$.

9.2 Local Coordinate Descriptions of an Affine Control System

The previous section presented a construction for the Grassmann bundle $G_1^{[0,\Delta]}(M)$ which models an affine control system. A part of this construction involved the set of charts Λ which consisted of all triples $((x,t),F,U_x)$ where $(x,t) : U_x \to \mathcal{R}^{(n+1)}$ is a local coordinate chart for $M\times\mathcal{R}$ and F is a partial frame which is defined at each point $p \in U_x$ and which satisfies the requirement that $\text{span}\{F(p)\} = \Delta(p)$ at each point $p \in \Delta$. In this section, we will examine how the charts in Λ are related to the vector field description of an affine control system and to the notion of feedback equivalence. The previous section also discussed the fact that the bundle $G_1^{[0,\Delta]}(M)$ could be endowed with a set of local coordinate charts which turn it into a differentiable manifold. In this section we will also examine what a neighborhood of the manifold $G_1^{[0,\Delta]}(M)$ looks like with respect to these local coordinate charts.

We begin by recalling some concepts from the vector field approach. An affine control system with m inputs $u^1, \ldots, u^m$ consists of a state space manifold M together with a collection of smooth vector fields $f, g_1, \ldots, g_m$ defined over M. Over a local coordinate chart $x : U_x \subset M \rightarrow \mathcal{R}^n$, the control system can be described by a set of differential equations of form

$$\dot{x} = \tilde{f}(x) + \sum_{i=1}^{m} \tilde{g}_i(x) u^i \tag{9.2}$$

where the u^i are scalar variables and the local description of each vector field has the form

$$\tilde{g}_i(x) := \sum_{j=1}^{n} g_i^j \circ x^{-1}(x) \frac{\partial}{\partial x^j}.$$

At each point x, the variables u^i parameterize an affine subset of the tangent space $T_x(x(U_x))$. If (y, U_y) is another coordinate system, the control system can also be described over U_y by the equations

$$\dot{y} = \hat{f}(y) + \sum_{i=1}^{m} \hat{g}_i(y) v^i. \tag{9.3}$$

If $U_x \cap U_y \neq \emptyset$, then there is a local coordinate transformation $y \circ x^{-1} : x(U_x) \rightarrow y(U_y)$ which we will denote by $y(x)$. On $U_x \cap U_y$, the two local descriptions of the control system must satisfy the relationships

$$\begin{aligned} \hat{f}(y(x)) &= \frac{\partial y}{\partial x} \tilde{f}(x) \\ \hat{g}_i(y(x)) &= \frac{\partial y}{\partial x} \tilde{g}_i(x). \end{aligned} \tag{9.4}$$

Any two local descriptions of a control system which satisfy equation 9.4 are said to be equivalent with respect to state transformation. If we were to form the disjoint union of all such local descriptions and then form equivalence classes using equation 9.4 as our equivalence relation, we would produce a geometric object which would be isomorphic to our original description of the system as a smooth manifold together with a collection of vector fields. Therefore, this description of an affine control system really contains the notion of equivalence under state transformation as a part of its structure.

We can also define the two local representations of a control system defined by equations 9.2 and 9.3 to be equivalent if and only if they satisfy the relations

$$\begin{aligned} \frac{\partial y}{\partial x} \tilde{f}(x) &= \left(\hat{f}(y(x)) + \sum_{k=1}^{m} \hat{g}_k(y(x)) \alpha^k(x) \right) \\ \frac{\partial y}{\partial x} \tilde{g}_i(x) &= \left(\sum_{k=1}^{m} \hat{g}_k(y(x)) \beta_i^k(x) \right) \end{aligned} \tag{9.5}$$

for some set of smooth functions $\alpha^k(x)$, $\beta_i^k(x)$ defined over $x(U_x \cap U_y)$. Any pair of local representations of a control system which satisfy the relations 9.5 are said to be equivalent with respect to state feedback transformations. Thus, two local representations of a control system are equivalent with respect to state feedback if there exists a change of coordinates on the state space together with a state feedback such that one representation can be transformed into the other. We are going to show that this notion of equivalence is built into the structure of the bundle $G_1^{[0,\Delta]}(M \times \mathcal{R})$. To show this, we begin by extending each of the local descriptions 9.2 and 9.3 to $U_x \times \mathcal{R}$ and $U_y \times \mathcal{R}$ by appending the dynamics $\dot{t} = 1$ to each system. The vector fields associated with these extended systems can be viewed as two matrices. Over $U_x \times \mathcal{R}$, we have the matrix

$$\tilde{F} := \begin{bmatrix} \tilde{g}_1 & \cdots & \tilde{g}_m & \tilde{f} \\ 0 & & 0 & 1 \end{bmatrix}$$

which is defined relative to the basis

$$\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial t} \right\},$$

and over $U_y \times \mathcal{R}$, we have the matrix

$$\hat{F} := \begin{bmatrix} \hat{g}_1 & \cdots & \hat{g}_m & \hat{f} \\ 0 & & 0 & 1 \end{bmatrix}$$

which is defined relative to the basis

$$\left\{ \frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n}, \frac{\partial}{\partial t} \right\}.$$

In matrix form, the equivalence relation 9.5 can be written as

$$\begin{bmatrix} \hat{g}_1 & \cdots & \hat{g}_m & \hat{f} \\ 0 & & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta & \alpha \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\partial y}{\partial x} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{g}_1 & \cdots & \tilde{g}_m & \tilde{f} \\ 0 & & 0 & 1 \end{bmatrix}. \tag{9.6}$$

From this equation, it is apparent that the two local representations of the control system are equivalent with respect to state feedback if and only if their extensions span the same distribution over the state-time manifold. Thus, the collection of all local representations of an affine system which are equivalent with respect to state feedback define a distribution over $M \times \mathcal{R}$, and the triples $((x,t), \tilde{F}, U_x)$ and $((y,t), \hat{F}, U_y)$ are elements of the collection of frames, Λ, which is associated with this distribution. If we denote the distribution defined by these local representations as Δ, then we can generate the Grassmann bundle $G_1^{[0,\Delta]}(M)$ using the construction described in the previous section. We can also follow the procedure described in the previous section to generate an atlas of local coordinate charts on $G_1^{[0,\Delta]}(M)$ which give it the structure of a differentiable manifold. Associated with each triple $((x,t), \tilde{F}, U_x)$, we have a local trivialization

$$t_{((x,t),\tilde{F},U_x)} : \pi^{-1}(U_x) \to U_x \times G_1^{m+1}.$$

On $U_x \times G_1^{m+1}$, we will consider the open subset V which contains all points of the form

$$((p,t), l) = ((p,t), \text{span}\{[a^1 \cdots a^m \, a^t]^T\})$$

with $a^t \neq 0$. On V, we can define a coordinate chart $c : V \to \mathcal{R}^{n+m+1}$ which maps any point $((p,t), \text{span}\{[a^1 \cdots a^m a^t]^T\}) \in V$ to the point $(x(p), t, a^1/a^t, \ldots, a^m/a^t) \in \mathcal{R}^{n+m+1}$. By taking the composition of c with the local trivialization $t_{((x,t),\tilde{F},U_x)}$, we can form a coordinate chart for $G_1^{[0,\Delta]}(M)$ over the open set defined by $t^{-1}_{((x,t),\tilde{F},U_x)}(V)$. On the intersection of any two such charts, we have a local coordinate transformation

$$c \circ t_{((y,t),\hat{F},U_y)} \circ t^{-1}_{((x,t),\tilde{F},U_x)} \circ c^{-1}.$$

To see what this looks like, we will first consider the transformation

$$t_{((y,t),\hat{F},U_y)} \circ t^{-1}_{((x,t),\tilde{F},U_x)} : U_x \cap U_y \times G_1^{m+1} \to U_x \cap U_y \times G_1^{m+1}$$

which can be described pointwise by the equation

$$((p,t), \mathrm{span}\{[u^1 \cdots u^m \ 1]^T\}) \to ((p,t), \Phi(T)(\mathrm{span}\{[u^1 \cdots u^m \ 1]^T\})).$$

The transformation T in this equation is defined by the relation $\hat{F}T = \frac{\partial y}{\partial x}\tilde{F}$. We are assuming that the charts defined by equations 9.2 and 9.3 satisfy equation 9.6, so we must have

$$T = \begin{bmatrix} \beta & \alpha \\ 0 & 1 \end{bmatrix}.$$

Consequently, the mapping $\Phi(T)(\mathrm{span}\{[u^1 \cdots u^m \ 1]^T\})$ will have the form

$$\begin{bmatrix} u \\ 1 \end{bmatrix} \to \begin{bmatrix} \beta & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ 1 \end{bmatrix},$$

and the coordinate transformation $c \circ t_{((y,t),\hat{F},U_y)} \circ t^{-1}_{((x,t),\tilde{F},U_x)} \circ c^{-1}$ can be described pointwise by the equation

$$(x,t,u) \to (y(x), t, \alpha(x) + \sum_{k=1}^{m} \beta_m(x) u^m).$$

9.3 Prolongation and Dynamic State Feedback

In this section, we will show how the Grassmann bundle model described in the previous sections can be extended to encompass a class of dynamic state feedbacks. We will begin by giving a general definition of dynamic state feedback, and then we will describe the class of dynamic state feedbacks which we will be considering.

9.3.1 Dynamic State Feedback

Definition 9.3.1 *Given a control system which is locally described by the equation*

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u^i, \tag{9.7}$$

an affine dynamic compensator or dynamic state feedback is defined as a system

$$\begin{aligned} \dot{z} &= \phi(z,x) + \textstyle\sum_{i=1}^{n} \gamma_i(z,x) v^i \\ u &= \alpha(z,x) + \textstyle\sum_{i=1}^{n} \beta_i(z,x) v^i \end{aligned}$$

where $z \in \mathcal{R}^q$, $\phi : \mathcal{R}^q \times M \to \mathcal{R}^q$, $\gamma : \mathcal{R}^q \times M \to \mathcal{R}^{q\times m}$, $\alpha : \mathcal{R}^q \times M \to \mathcal{R}^m$, and $\beta : \mathcal{R}^q \times M \to \mathcal{R}^{m\times m}$ and $v \in \mathcal{R}^m$ is a new control input.

The compensated system is obtained by interconnecting the compensator output with the inputs of system 9.7 to obtain

$$\begin{aligned} \dot{x} &= \left(f(x) + \sum_{i=1}^{m} g_i(x) \alpha^i(z,x) \right) + \sum_{j=1}^{n} \left(\sum_{i=1}^{m} g_i(x) \beta_j^i(z,x) \right) v^j \\ \dot{z} &= \phi(z,x) + \sum_{i=1}^{n} \gamma_i(z,x) v^i. \end{aligned}$$

In the development which follows, we will consider the class of dynamic compensators which have the form

$$\begin{aligned} \dot{z}^i &= v^{m-r+i} \\ u &= \left(\alpha(x) + \sum_{k=1}^{r} \beta_{k+m-r}(z,x) z^k\right) + \sum_{j=1}^{m-r} \beta_j(z,x) v^j \end{aligned} \tag{9.8}$$

where $r \leq m$ and $1 \leq i \leq r$. Any compensator of this type can be generated by first applying a static state feedback to system 9.7 and then appending integrators to the last r inputs. By iterating this construction, we can also produce more complex dynamic compensators.

9.3.2 The Canonical Prolongation of a Control System

We are now going to show how a feedback compensator of the type 9.8 can be constructed geometrically using the Grassmann bundle model of the control system.

We will first consider the case where $r = m$. Physically, this corresponds to adding an integrator to each control input. We will assume that we have constructed the Grassmann bundle $G_1^{[0,\Delta]}(M \times \mathcal{R})$ which represents the control system 9.7. As part of the bundle structure, we have the projection map $\pi : G_1^{[0,\Delta]}(M \times \mathcal{R}) \to M \times \mathcal{R}$ which maps a point in the bundle to its corresponding base point on the state-time space $((p,t),l) \to (p,t)$. We can use the pullback map, π^*, to define a codistribution Ω over $G_1^{[0,\Delta]}(M \times \mathcal{R})$ which is defined pointwise by the equation

$$\Omega_{((p,t),l)} = \pi^*(l^{\perp})|_{(p,t)}.$$

Since $l_{(p,t)}$ is a one-dimensional subspace of an $(n+1)$-dimensional tangent space, $(l^{\perp})|_{(p,t)}$ is a n-dimensional subspace of the cotangent space and Ω is an n-dimensional codistribution. The Grassmann bundle $G_1^{[0,\Delta]}(M \times \mathcal{R})$ is an $(n+m+1)$-dimensional manifold, so the distribution $\Omega^{\perp}$ must be $(m+1)$-dimensional. We can use the distribution $\Omega^{\perp}$ to construct a new Grassmann bundle over $G_1^{[0,\Delta]}(M \times \mathcal{R})$. We will denote this bundle by $G_1^{[0,\Omega^{\perp}]}\left(G_1^{[0,\Delta]}(M \times \mathcal{R})\right)$. It is not hard to see that if we repeat this process again, we will obtain another $(m+1)$-dimensional distribution which is defined over a $(2m+n+1)$-dimensional space, and that if we iterate the process p times, we will obtain an $(m+1)$-dimensional distribution which is defined over a $(pm+n+1)$-dimensional manifold.

To see what this process looks like with respect to a local coordinate chart on $G_1^{[0,\Delta]}(M \times \mathcal{R})$, we restrict our attention to an open neighborhood $V \subset G_1^{[0,\Delta]}(M \times \mathcal{R})$ on which we have the local coordinates x , t, and u. Each point (x,t,u) defines a line

$$\operatorname{span}\left\{\begin{bmatrix} f(x,t) + \sum_{j=1}^{m} g_j(x,t)u^j \\ 1 \end{bmatrix}\right\}$$

in the tangent space $T_{(p,t)}(M \times \mathcal{R})$ and a subspace $l^{\perp}_{(x,t)}$ of $T^*(p,t)(M \times \mathcal{R})$ which is spanned by the covectors

$$\{dx^1 - (f^1(x,t) + \sum_{j=1}^{m} g_j^1(x,t)u^j)dt, \ldots, dx^n - (f^n(x,t) + \sum_{j=1}^{m} g_j^n(x,t)u^j)dt\}.$$

Pulled back to V, these one-forms locally define the codistribution Ω. The distribution $\Omega^{\perp}$ will be spanned by the tangent vectors

$$\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^m}, \sum_{i=1}^{n}(f^i(x,t) + \sum_{j=1}^{m} g_j^i(x,t)u^j)\frac{\partial}{\partial x^i} + 1\frac{\partial}{\partial t}.$$

Written in matrix form, this distribution looks like

$$\Omega^{\perp} = \operatorname{span}\begin{bmatrix} 0 & f + \sum_{j=1}^{m} g_j(x,t)u^j \\ I_{m\times m} & 0 \\ 0 & 1 \end{bmatrix}.$$

If we were to iterate this process p times, we would obtain a distribution of the form

$$\Omega_p^\perp = \text{span}\begin{bmatrix} 0 & f+\sum_{j=1}^m g_j(x,t)u^j \\ 0 & \dot{u} \\ & \vdots \\ 0 & \overset{(p)}{u} \\ I_{m\times m} & 0 \\ 0 & 1 \end{bmatrix}$$

which represents a system with p integrators appended to each input.

9.3.3 General Prolongations of a Control System

We now turn our attention to the more general case of dynamic state feedbacks of the type 9.8 in which $r < m$. In this more general case, the prolongation process is basically the same, but in order to describe it, we first need to define a new Grassmann bundle. Suppose that we have an affine control system of the form 9.7 and we want to append integrators to the last r inputs. The vectors $g_1,\ldots,g_{m-r}$ which correspond to the unaffected inputs span an $(m-r)$-dimensional distribution Δ_1. This distribution forms a filtration $\Delta_1 \subset \Delta$ over $M\times\mathcal{R}$. We can use this filtration to construct a Grassmann bundle $G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})$ whose fibres over each point (p,t) consist of the set of all $(m-r+1)$-dimensional subspaces of $T_{(p,t)}(M\times\mathcal{R})$ which contain the subspace $\Delta_1(p,t)$ and are contained in the subspace $\Delta(p,t)$. Recall that in Chapter 5 we showed that if Δ_1 is a p-dimensional subspace, Δ is a r-dimensional subspace, and $\Delta_1\subset\Delta$, then the Grassmann manifold $G^{[\Delta_1,\Delta]}_{p+k}$ is diffeomorphic to the Grassmann manifold G^{r-p}_k. Using this fact, we find that each fibre of $G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})$ is diffeomorphic to the Grassmann manifold G^{r+1}_1.

In order to construct this bundle, we form the collection Λ of all quadruples of the form $((x,t),F_1,F_2,U_x)$ where (x,t) is a local coordinate chart defined over U_x, F_1 is a local frame for the distribution Δ_1 which is defined over U_x, and F_2 is an additional set of r linearly independent vector fields such that F_1 and F_2 taken together span Δ. We will also assume that $dt|_{\Delta_1}\equiv 0$ and $dt|_\Delta \neq 0$. If the domains of two charts $((y,t),F_1,F_2,U_y)$ and $((x,t),G_1,G_2,U_x)$ intersect, then on $U_x\cap U_y$ we must have that

$$\begin{bmatrix} F_1 & F_2 \end{bmatrix}\begin{bmatrix} M_1^1 & M_2^1 \\ 0 & M_2^2 \end{bmatrix} = \begin{bmatrix} \frac{\partial y}{\partial x} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} G_1 & G_2 \end{bmatrix}$$

for some set of smooth matrices M_1^1, M_2^1, and M_2^2. Following the procedure introduced in Chapter 7, we form the disjoint union

$$\mathcal{U} = \bigcup_{((x,t),F_1,F_2,U_x)\in\Lambda} \{((x,t),F_1,F_2,U_x)\}\times U_{((x,t),F_1,F_2,U_x)}\times G_1^{r+1}$$

and then define equivalent classes on $\mathcal{U}$ using the equivalence relation

$$(((y,t),F_1,F_2,U_y),(p,t),l_y) \sim (((x,t),G_1,G_2,U_x),(q,\tau),l_x) \tag{9.9}$$

if and only if $(p,t)=(q,\tau)$ and $l_y=\Phi(M_2^2(p,t))l_x$. The set of all equivalence classes endowed with the quotient topology forms the total space of the bundle $G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})$.

We can now form a codistribution Ω over $G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})$ which is defined pointwise by $\Omega(p,t,D)=\pi^* D^\perp_{(p,t)}$ where D is a $(m-r+1)$-dimensional subspace of $T_{(p,t)}(M\times\mathcal{R})$ and $D^\perp$ is a $(n-m+r)$-dimensional subspace of $T^*_{(p,t)}(M\times\mathcal{R})$. The manifold $G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})$ is $(n+r+1)$-dimensional, so the distribution $\Omega^\perp$ has dimension $(m+1)$. This distribution can be used to construct a new Grassmann bundle $G_1^{[0,\Omega^\perp]}\left(G^{[\Delta_1,\Delta]}_{(m-r+1)}(M\times\mathcal{R})\right)$ which models the dynamically compensated system.

9.4 Nonaffine Systems

In this section, we will consider systems which can be described locally by equations of the form

$$\dot{x} = f(x, u, t). \tag{9.10}$$

Equations of this form can also be modeled as Grassmann bundles, but some modifications to the theory are required. The Grassmann bundle $G_1^{[0,\Delta]}(M \times \mathcal{R})$ associated with an affine control system can be viewed as a subbundle of the bundle $G_1^{(n+1)}(M \times \mathcal{R})$ whose fibres over a point (p, t) consist of the collection of all one-dimensional subspaces of the tangent space $T_{(p,t)}(M \times \mathcal{R})$. Because of the affine structure, the fibres of $G_1^{(n+1)}(M \times \mathcal{R})$ over the point (p, t) span a $(m + 1)$-dimensional subspace of $T_{(p,t)}(M \times \mathcal{R})$. We can view a nonaffine system as a more general subbundle of $G_1^{(n+1)}(M \times \mathcal{R})$ whose fibres over (x, t) are diffeomorphic a to more general submanifold of the Grassmann manifold $G_1^{(n+1)}$. In equation 9.10, the function $f(x, u, t)$ can be viewed as a local parameterization of this submanifold. In order to study such systems, we will start with the first prolongation of the bundle $G_1^{(n+1)}(M \times \mathcal{R})$ which is defined by the codistribution Ω defined pointwise for each $(p, t, l) \in G_1^{(n+1)}(M \times \mathcal{R})$ by

$$\Omega(p, t, l) = \pi^{-1} l^{\perp}_{(p,t)}.$$

This codistribution can be locally described by the span of the one-forms

$$\Omega = \text{span}\{dx^1 - \dot{x}^1 dt, \ldots, dx^n - \dot{x}^n dt\}.$$

Restricted to the submanifold defined by $\dot{x} = f(x, u, t)$, these forms become

$$\Omega = \text{span}\{dx^1 - f^1(x, u, t)dt, \ldots, dx^n - f^n(x, u, t)dt\}.$$

Note that if the control system is affine, then we obtain the forms

$$\Omega = \{dx^1 - (f^1(x, t) + \sum_{j=1}^{m} g_j^1(x, t)u^j)dt, \ldots, dx^n - (f^n(x, t) + \sum_{j=1}^{m} g_j^n(x, t)u^j)dt\}.$$

In the last section, we showed that these forms locally span the codistribution associated with the first canonical prolongation of the affine system. We can use this observation to consider the problem of whether or not there exists a change of coordinates $(x, t, u) \to (y, t, v)$ which turns 9.10 into an affine system.

Theorem 9.4.1 *Let $\dot{x} = f(x, u, t)$ be a local representation of a nonaffine control system which can be modeled by a codistribution*

$$\Omega = span\{dx^1 - f^1(x, u, t)dt, \ldots, dx^n - f^n(x, u, t)dt\} \tag{9.11}$$

defined over an open subset of $\mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}$. There exists a coordinate transformation $(x, u, t) \to (z(x, t), v(x, u, t), t)$ and a set of one-forms of the type

$$dz^i - (f^i(z, t) + \sum_{j=1}^{m} g_j^i(z, t)v^j)dt$$

which satisfies the relation

$$\Omega = \{dz^1 - (f^1(z, t) + \sum_{j=1}^{m} g_j^1(z, t)v^j)dt, \ldots, dx^n - (f^n(z, t) + \sum_{j=1}^{m} g_j^n(z, t)v^j)dt\} \tag{9.12}$$

if and only if the retracting space of the codistribution defined by

$$\hat{\Omega} := \{\omega \in \Omega | d\omega = 0 \ mod \ \Omega\}$$

satisfies $C(\hat{\Omega}) \subset span \ \{dx^1, \ldots, dx^n, dt\}$.

Proof: If the exterior differential system 9.11 is feedback equivalent to an affine system, then we can work relative to the exterior differential system 9.12. To compute $\hat{\Omega}$, we will take the exterior derivative of an arbitrary element

$$\alpha = \sum_{i=1}^{n} a_i(z,v,t)(dz^i - (f^i(z,t) + \sum_{j=1}^{m} g_j^i(z,t)v^j)dt) \in \Omega$$

and find the conditions under which it satisfies the condition

$$d\alpha = 0 \bmod \Omega. \tag{9.13}$$

The 2-form $d\alpha$ can be written with respect to the local coordinates as

$$d\alpha = \sum_{i=1}^{n} da_i(z,v,t) \wedge (dz^i - (f^i(z,t) + \sum_{j=1}^{m} g_j^i(z,t)v^j)dt)$$

$$+ \sum_{i=1}^{n} -a_i(z,v,t)(df^i(z,t) + \sum_{j=1}^{m} dg_j^i(z,t)v^j + \sum_{j=1}^{m} g_j^i(z,t)dv^j) \wedge dt$$

$$\Rightarrow d\alpha = -\sum_{i=1}^{n}\sum_{j=1}^{m} a_i(z,v,t)g_j^i(z,t)dv^j \wedge dt \bmod \Omega.$$

The coefficients a_i will satisfy this condition 9.13 if and only if they are in the left null space of the $n \times m$ matrix

$$\begin{bmatrix} g_1 & \cdots & g_m \end{bmatrix}. \tag{9.14}$$

Since we are assuming that the vectors g_i are linearly independent, the left null space of the matrix 9.14 will be $(n-m)$-dimensional. Therefore, there will be $(n-m)$ linearly independent one-forms α^k, $1 \leq k \leq n-m$, which span $\hat{\Omega}$. Furthermore, since the coefficients of the g_i's are only functions of the z and t coordinates, there exists coefficients a_i^k which are also only functions of z and t. Consequently, each α^k takes the form

$$\alpha^k = \sum_{i=1}^{n} a_i^k(z,t)(dz^i - (f^i(z,t) + \sum_{j=1}^{m} g_j^i(z,t)v^j)dt)$$

$$\Rightarrow \alpha^k = \sum_{i=1}^{n} a_i^k(z,t)dz^i - (\sum_{i=1}^{n} a_i^k(z,t)f^i(z,t))dt$$

From this, it is clear that the generators of $\hat{\Omega}$ can be expressed solely in terms of the coordinates z and t. Therefore, we must have $C(\hat{\Omega}) \subset \text{span}\{dz^1, \ldots, dz^n, dt\}$. Finally, the feedback transformation takes $(x,t) \to (z(x,t),t)$ so

$$dz^i = \sum_{j=1}^{n} \frac{\partial z^i}{\partial x^j} dx^j + \frac{\partial z^i}{\partial t} dt$$

and this implies that

$$\text{span}\{dz^1, \ldots, dz^n, dt\} = \text{span}\{dx^1, \ldots, dx^n, dt\}.$$

Consequently, $C(\hat{\Omega}) \subset \text{span}\{dx^1, \ldots, dx^n, dt\}$.

To prove the sufficiency of the condition that $C(\hat{\Omega}) \subset \text{span}\{dx^1, \ldots, dx^n, dt\}$ in Theorem 9.4.1, we begin by computing the distribution $\hat{\Omega}^\perp$ using the exterior differential system 9.11. Taking the exterior derivative of an arbitrary 1-form

$$\alpha = \sum_{i=1}^{n} a_i(x,u,t)(dx^i - f^i(x,u,t)dt) \in \Omega,$$

we find that

$$d\alpha = \sum_{i=1}^{n} da_i(x,u,t) \wedge (dx^i - f^i(x,u,t)dt)$$

$$+ \sum_{i=1}^{n} -a_i(x,u,t)df^i(x,u,t) \wedge dt$$

$$= \sum_{i=1}^{n} da_i(x,u,t) \wedge \alpha^i - \sum_{i=1}^{n}\sum_{j=1}^{n} a_i(x,u,t)\frac{\partial f^i}{\partial x^j}dx^j \wedge dt - \sum_{i=1}^{n}\sum_{k=1}^{m} a_i(x,u,t)\frac{\partial f^i}{\partial u^k}du^k \wedge dt$$

$$\Rightarrow d\alpha = -\sum_{i=1}^{n}\sum_{k=1}^{m} a_i(x,u,t)\frac{\partial f^i}{\partial u^k}du^k \wedge dt \text{ mod } \Omega.$$

As before, the coefficients $a_i(x,u,t)$ will only satisfy the condition 9.13 if they lie in the left null space of the $n \times m$ matrix $\frac{\partial f}{\partial u}$. Assuming that this matrix has full rank, its left null space will be $(n-m)$-dimensional, so there will be $(n-m)$ linearly independent one-forms α^k, $1 \leq k \leq n-m$, which span $\hat{\Omega}$.

Suppose that $C(\hat{\Omega}) = \text{span }\{dy^1, \ldots, dy^p\}$. Using Theorem 8.3.2, we know that we can select a set of one-forms

$$\omega^k = a_i^k(y)dy^i, \quad k = 1, \ldots, (n-m)$$

which span $\hat{\Omega}$ and whose coefficients are only functions of the y coordinates. By assumption, we also know that $C(\hat{\Omega}) \subset \text{span }\{dx^1, \ldots, dx^n, dt\}$, so there must exist p linearly independent functions $y^i(x,t)$ which satisfy $dy^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j}dx^j + \frac{\partial y^i}{\partial t}dt$. Consequently, the coefficients a_i^k can also be expressed as functions of x and t.

We can form a matrix

$$\begin{bmatrix} A_1^1(x,t) & A_2^1(x,t) & 0 & A_t^1(x,t) \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is defined relative to the cobasis $\{dx^1, \ldots, dx^n, du^1, \ldots, du^k, dt\}$. The first $(n-m)$ rows of this matrix span $\hat{\Omega}$. Inverting this matrix produces the matrix

$$\begin{bmatrix} (A_1^1)^{-1} & -(A_1^1)^{-1}A_2^1 & 0 & -(A_1^1)^{-1}A_t^1 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is defined relative to the basis $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^k}, \frac{\partial}{\partial t}\}$. The last $(2m+1)$ columns of this matrix span $\hat{\Omega}^\perp$. The distribution $\hat{\Omega}^\perp$ is also spanned by the last $(2m+1)$ columns of the matrix

$$\begin{bmatrix} I & \frac{\partial f^1}{\partial u} & 0 & f^1 \\ 0 & \frac{\partial f^2}{\partial u} & 0 & f^2 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is defined relative to the same basis. Therefore, there must exist a smooth $(2m+1) \times (2m+1)$ matrix of function which satisfies

$$\begin{bmatrix} -(A_1^1)^{-1}A_2^1 & 0 & -(A_1^1)^{-1}A_t^1 \\ I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1}{\partial u} & 0 & f^1 \\ \frac{\partial f^2}{\partial u} & 0 & f^2 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_1^1 & T_2^1 & T_3^1 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is not hard to see that the elements of this transformation matrix must satisfy $T_1^1 = (\frac{\partial f^2}{\partial u})^{-1}$, $T_2^1 = 0$, and $T_3^1 = -(\frac{\partial f^2}{\partial u})^{-1} f^2$. Consequently, we must also have $-(A_1^1)^{-1} A_2^1(x,t) = \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1}$ and $-(A_1^1)^{-1} A_t^1(x,t) = f^1 - \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} f^2$. Since the left-hand sides of these equations are functions of x and t, the same must be true for the right-hand sides. Consequently, the $(n-m)$ forms represented by the matrix

$$\begin{bmatrix} \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} & 0 & f^1 - \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} f^2 \end{bmatrix}$$

can be generated as the pull-backs of a set of forms $\hat{\alpha}^1, \ldots, \hat{\alpha}^{n-m}$ defined over $\mathcal{R}^n \times \mathcal{R}$.

Taking the first prolongation of this system gives a system defined over $G_1^{[0,\hat{\Omega}^\perp]}(\mathcal{R}^n \times \mathcal{R})$. This system can be locally described with respect to the coordinates x, t, and v by the matrix

$$\Delta_e = \begin{bmatrix} 0 & \left(\frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} v + f^1 - \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} f^2\right) \\ 0 & v \\ I & 0 \\ 0 & 1 \end{bmatrix} \tag{9.15}$$

which is defined relative to the basis $\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial v^1}, \ldots, \frac{\partial}{\partial v^k}, \frac{\partial}{\partial t}\}$. In order to complete the proof, we need to show that there exists a local diffeomorphism ϕ which maps an open subset of $G_1^{[0,\hat{\Omega}^\perp]}(\mathcal{R}^n \times \mathcal{R})$ onto $\mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}$ in such a way that $\phi_* \Delta_e = \Omega^\perp$. As a candidate transformation, consider the mapping ϕ defined such that $\phi^{-1} : (x,u,t) \to (x, f^2(x,u,t), t)$. We are assuming that $\frac{\partial f^2}{\partial u}$ is nonsingular, so locally there exists a function $g : (x,v,t) \to u$ which for fixed x and t satisfies $v = (f^2 \circ g)_{(x,t)}(v)$. Using the function g, the mapping ϕ can be described by $\phi : (x,v,t) \to (x, g(x,v,t), t)$. Pushing the distribution Δ_e forward using ϕ_*, we get

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ \frac{\partial g}{\partial x^1} & \frac{\partial g}{\partial x^2} & \frac{\partial g}{\partial u} & \frac{\partial g}{\partial t} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & \left(\frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} f^2) + f^1 - \frac{\partial f^1}{\partial u}(\frac{\partial f^2}{\partial u})^{-1} f^2\right) \\ 0 & f^2(x,u,t) \\ I & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & f^1 \\ 0 & f^2 \\ \frac{\partial g}{\partial u} & \left(\frac{\partial g}{\partial x} f + \frac{\partial g}{\partial t}\right) \\ 0 & 1 \end{bmatrix}$$

Finally, applying the feedback

$$\begin{bmatrix} \dot{u} \\ 1 \end{bmatrix} = \begin{bmatrix} (\frac{\partial g}{\partial u})^{-1} & (\frac{\partial g}{\partial u})^{-1}\left(\frac{\partial g}{\partial x} f + \frac{\partial g}{\partial t}\right) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\hat{u}} \\ 1 \end{bmatrix}$$

produces the desired transformation. □

Chapter 10

Invariance

This chapter is devoted to the development of several topics related to the invariance properties of distributions and codistributions. This material is useful for studying questions related to controllability, observability, disturbance decoupling, and noninteracting controls.

10.1 Invariance of A Distribution with Respect to Vector Fields

A distribution Δ defined over M is said to be invariant with respect to a vector field f if and only if $[f, \Delta] \subset \Delta$. Similarly, a codistribution Ω defined over M is said to be invariant with respect to a vector field f if and only if $L_f \Omega \subset \Omega$. The following lemma relates these two definitions.

Lemma 10.1.1 *A distribution Δ is invariant with respect to a vector field f if and only if $\Delta^\perp$ is also invariant with respect to the vector field f.*

Proof: If we select any vector field $v \in \Delta$ and any codistribution $\omega \in \Delta^\perp$, then, using equation 7.8 we can write

$$v \lrcorner L_f \omega = v \lrcorner f \lrcorner dw + v \lrcorner d(f \lrcorner \omega).$$

Using Cartan's formula, we also have

$$v \lrcorner f \lrcorner dw + v \lrcorner d(f \lrcorner \omega) = f \lrcorner d(v \lrcorner \omega) - [f, v] \lrcorner \omega.$$

Combining these equations, we get

$$v \lrcorner L_f \omega = -[f, v] \lrcorner \omega.$$

Since this holds for any $\omega \in \Delta^\perp$ and any $v \in \Delta$, we find that

$$\Delta \lrcorner L_f \Delta^\perp = [f, \Delta] \lrcorner \Delta^\perp.$$

The theorem follows from the following sequence of equivalent statements.

$$\begin{array}{rrcl} & L_f \Delta^\perp & \subset & \Delta^\perp \\ \Leftrightarrow & \Delta \lrcorner L_f \Delta^\perp & \equiv & 0 \\ \Leftrightarrow & [f, \Delta] \lrcorner \Delta^\perp & \equiv & 0 \\ \Leftrightarrow & \Delta^\perp & \subset & [f, \Delta]^\perp \\ \Leftrightarrow & [f, \Delta] & \subset & \Delta. \end{array}$$

□

The next lemma presents some useful closure properties of invariant distributions.

Lemma 10.1.2 *If two distributions Δ_1 and Δ_2 are invariant with respect to a vector field f, then the following conditions also hold.*

1. $[f, \Delta_1 + \Delta_2] \subset \Delta_1 + \Delta_2$.
2. $[f, \Delta_1 \cap \Delta_2] \subset \Delta_1 \cap \Delta_2$.
3. $[f, [\Delta_1, \Delta_1]] \subset [\Delta_1, \Delta_1]$.
4. $[f, A(\Delta^\perp)] \subset A(\Delta^\perp)$, *where $A(\Delta^\perp)$ denotes the Cauchy characteristic associated with $\Delta^\perp$.*

Proof: Since Δ_1 and Δ_2 are assumed invariant with respect to f, the equations $[f, \Delta_1] \subset \Delta_1$ and $[f, \Delta_2] \subset \Delta_2$ must hold. Using these two facts, we can prove statements (1) and (2) by direct computation.

$$\begin{array}{lcccc} (1) & [f, \Delta_1 + \Delta_2] & \subset & [f, \Delta_1] + \Delta_1 + [f, \Delta_2] + \Delta_2 & \subset & \Delta_1 + \Delta_2. \\ (2) & [f, \Delta_1 \cap \Delta_2] & \subset & [f, \Delta_1] \cap [f, \Delta_2] & \subset & \Delta_1 \cap \Delta_2. \end{array}$$

To prove statement (3), we note that the invariance of Δ_1 implies that

$$[[f, \Delta_1], \Delta_1] \quad \subset \quad [\Delta_1, \Delta_1].$$

In particular, for any vector fields $v_1, v_2 \in \Delta_1$, we must have

$$\begin{aligned} [[f, v_1], v_2] &\subset [\Delta_1, \Delta_1] \\ [[f, v_2], v_1] &\subset [\Delta_1, \Delta_1] \\ \Rightarrow [[f, v_1], v_2] - [[f, v_2], v_1] &\subset [\Delta_1, \Delta_1]. \end{aligned} \tag{10.1}$$

The vectors v_1, v_2 and f are also related through the Jacobi identity

$$[[f, v_1]v_2] - [[f, v_2], v_1] = [f, [v_1, v_2]].$$

Substituting the right side of this equation into equation 10.1 gives

$$[f, [v_1, v_2]] \quad \subset \quad [\Delta_1, \Delta_1].$$

Since v_1 and v_2 were arbitrary vector fields in Δ_1, the result follows. Finally, to prove (4) we will use equation 7.9. For any $v \in A(\Delta^\perp)$ and any $\omega \in \Delta^\perp$, applying equation 7.9 to $d\omega$ gives

$$\begin{aligned} L_f(v \lrcorner d\omega) - v \lrcorner L_f d\omega &= [f, v] \lrcorner d\omega \\ \Leftrightarrow L_f(v \lrcorner d\omega) - v \lrcorner dL_f\omega &= [f, v] \lrcorner d\omega. \end{aligned} \tag{10.2}$$

Since $v \in A(\Delta^\perp)$, and Δ is invariant with respect to f, both of the terms on the left-hand side of equation 10.2 are in $\Delta^\perp$, so this equation implies that $[f, v] \lrcorner d\omega \in \Delta^\perp$. Similarly, applying equation 7.9 to ω gives

$$L_f(v \lrcorner \omega) - v \lrcorner L_f\omega \quad = \quad [f, v] \lrcorner \omega. \tag{10.3}$$

Again, both the terms on the left-hand side are in $\Delta^\perp$, so this equation implies that $[f, v] \lrcorner \omega \in \Delta^\perp$. This is true for every $\omega \in \Delta^\perp$, so $[f, v] \in A(\Delta^\perp)$. Furthermore, this is also true for every $v \in A(\Delta^\perp)$, so we find that $[f, A(\Delta^\perp)] \subset A(\Delta^\perp)$, and the result follows. □

As a corollary to Lemma 10.1.2, we can state the analogous closure properties for codistributions.

Corollary 10.1.1 *If two codistributions Ω_1 and Ω_2 are invariant with respect to a vector field f, then the following conditions also hold.*

1. $L_f(\Omega_1 \cap \Omega_2) \subset \Omega_1 \cap \Omega_2$.
2. $L_f(\Omega_1 + \Omega_2) \subset \Omega_1 + \Omega_2$.
3. $L_f\hat{\Omega}_1 \subset \hat{\Omega}_1$ *where* $\hat{\Omega} := \{\omega \in \Omega | d\omega = 0 \ mod \ \Omega\}$.
4. $L_f C(\Delta^\perp) \subset C(\Delta^\perp)$ *where $C(\Delta^\perp)$ denotes the retracting space associated with $\Delta^\perp$.*

Proof: The proof follows by setting $\Omega_1 = \Delta_1^{\perp}$ and $\Omega_2 = \Delta_2^{\perp}$ and applying Lemma 10.1.1 to each of the conditions in Lemma 10.1.2. □

If a distribution or codistribution is invariant with respect to a vector field f, and the distribution or codistribution is also completely integrable, then there exists a coordinate transformation such that, in the transformed coordinates, the system assumes a particularly simple form.

Lemma 10.1.3 *Let Δ be a nonsingular and involutive distribution of dimension d and suppose that Δ is invariant under the vector field f. Then at each point $p \in M$ there exists a neighborhood U_p of p and a coordinate transformation $z = \Phi(x)$ defined on U_p in which the vector field f is represented by a vector of the form*

$$f(z) = \begin{bmatrix} f^1(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) \\ \vdots \\ f^d(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) \\ f^{d+1}(z^{d+1}, \ldots, z^n) \\ \vdots \\ f^n(z^{d+1}, \ldots, z^n) \end{bmatrix}. \tag{10.4}$$

Proof: Since Δ is completely integrable, there exists a set of coordinates $z^{d+1}, \ldots, z^n$ whose differentials span $\Delta^{\perp}$. Since Δ is invariant with respect to f, Lemma 10.1.1 implies that $\Delta^{\perp}$ is also invariant with respect to f, so, for each i in the range $d+1 \leq i \leq n$, the 1-form dz^i must satisfy $L_f dz^i \in \Delta^{\perp}$. Consequently,

$$\begin{aligned} L_f dz^i &= \textstyle\sum_{j=d+1}^{n} a_j^i(z) dz^j \\ \Rightarrow \quad dL_f z^i &= \textstyle\sum_{j=d+1}^{n} a_j^i(z) dz^j. \end{aligned}$$

With respect to the z coordinates, the vector field f can be written as

$$f := \sum_{i=1}^{n} f^i(z) \frac{\partial}{\partial z^i}.$$

so the above equation can be written as

$$\begin{aligned} dL_f z^i &= \textstyle\sum_{j=d+1}^{n} a_j^i(z) dz^j \\ \Rightarrow \quad df^i(z) &= \textstyle\sum_{j=d+1}^{n} a_j^i(z) dz^j. \end{aligned}$$

Finally, this implies that for each i in the range $d+1 \leq i \leq n$ and each j in the range $1 \leq j \leq d$

$$\frac{\partial f^i}{\partial z^j} = 0.$$

Consequently, the last $(n-d)$ coefficients of f are not functions of the coordinates $z^1, \ldots, z^d$. □

Corollary 10.1.2 *Let Ω be a nonsingular and involutive codistribution of dimension $(n-d)$ which is invariant under the vector field f. Then at each point $p \in M$ there exists a neighborhood U_p of p and a coordinate transformation $z = \Phi(x)$ defined on U_p such that the vector field f can represented in the form 10.4.*

Proof: Apply Lemma 10.1.3 to $\Omega^{\perp}$. □

Geometrically, Lemma 10.1.3 implies that the flow associated with the vector field f takes integral surfaces of Δ into integral surfaces. This is illustrated in Figure 10.1 below. The surfaces in this picture represent integral manifolds of a distribution Δ. The curves represent the flow corresponding

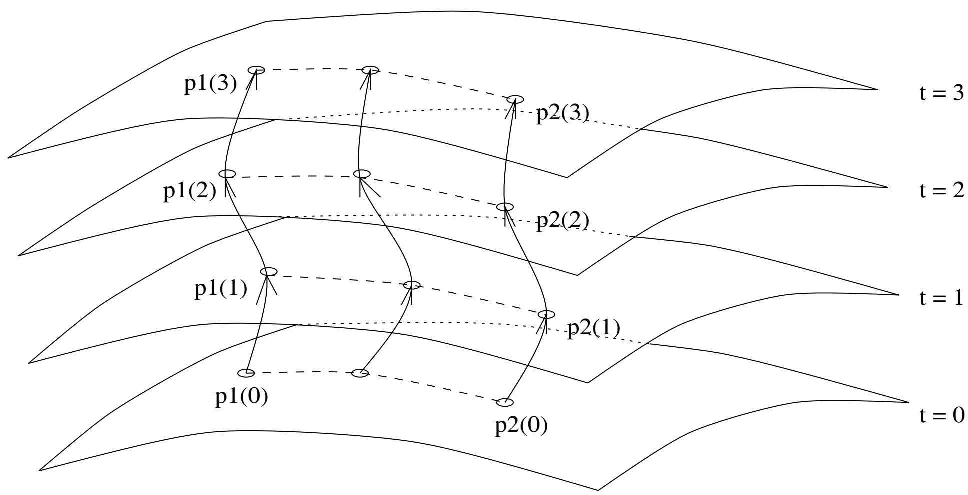

Figure 10.1: Flow of the Vector Field f Between Integral Submanifolds of Δ

to a vector field f. Under this flow, any two points p_1 and p_2 which begin on the same integral surface at time $t = 0$ are mapped to the same integral surface at every subsequent time.

If $f_1, \ldots, f_m$ is a collection of vector fields then a distribution Δ is said to be invariant with respect to the collection $\{f_0, \ldots, f_m\}$ if and only if

$$\begin{aligned} [f_0, \Delta] &\subset \Delta \\ [f_2, \Delta] &\subset \Delta \\ &\vdots \\ [f_m, \Delta] &\subset \Delta. \end{aligned}$$

If a completely integrable distribution or codistribution is invariant with respect to a collection of vector fields, then Lemma 10.1.3 and Corollary 10.1.2 imply that each vector field in the collection can be represented in the form 10.4. We will formalize this statement in the following corollary to Lemma 10.1.3.

Corollary 10.1.3 *Let Δ be a nonsingular and involutive distribution of dimension d and suppose that Δ is invariant under the set of vector fields $\{f_1, \ldots, f_m\}$. Then at each point $p \in M$ there exists a neighborhood U_p of p and a coordinate transformation $z = \Phi(x)$ defined on U_p on which the set can be represented as the matrix*

$$M(z) = \begin{bmatrix} f_1^1(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) & \cdots & f_m^1(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) \\ \vdots & & \vdots \\ f_1^d(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) & & f_m^d(z^1, \ldots z^d, z^{d+1}, \ldots, z^n) \\ f_1^{d+1}(z^{d+1}, \ldots, z^n) & & f_m^{d+1}(z^{d+1}, \ldots, z^n) \\ \vdots & & \vdots \\ f_1^n(z^{d+1}, \ldots, z^n) & \cdots & f_m^n(z^{d+1}, \ldots, z^n) \end{bmatrix}$$

which is defined relative to the basis $\{\frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^n}\}$.

We can exploit this fact to represent an affine control system in a form which isolates a set of states which are not affected by the control inputs.

Lemma 10.1.4 *Let Δ be a nonsingular and involutive distribution of dimension d and suppose that Δ is invariant under the vector fields*

$$f, g_1, \ldots, g_m.$$

Furthermore, assume that

$$g_1, \ldots, g_m \subset \Delta.$$

Then at each point $p \in M$ there exists a neighborhood U_p of p and a coordinate transformation $z = \Phi(x)$ defined on U_p such that the control systems is represented in the form

$$\begin{array}{rcll} \dot{z}^1 & = & f^1(z^1, z^2) & + \sum_{i=1}^m g_i^1(z^1, z^2)u^i \\ \dot{z}^2 & = & f^2(z^2). & \end{array}$$

Proof: The proof follows directly from Lemma 10.1.3 and the fact that the vector fields $g_1, \ldots, g_m$ are assumed to be contained in the distribution Δ. □

If an affine control system has p outputs $h^1(z), \ldots, h^p(z)$, Lemma 10.1.3 can also be used to represent the control system in a form which isolates a set of states which are indistinguishable from the outputs.

Lemma 10.1.5 *Let Δ be a nonsingular and involutive distribution of dimension d and suppose that Δ is invariant under the vector fields $f, g_1, \ldots,$ and g_m. Furthermore, assume that the system has linearly independent outputs $h^1(x), \ldots, h^p(x)$, and that*

$$\Delta \subset ker\{dh^1, \ldots, dh^p\}.$$

Then at each point $p \in M$ there exists a neighborhood U_p of p and a coordinate transformation $z = \Phi(x)$ defined on U_p such that the control system with outputs is represented in the form

$$\begin{array}{rcll} \dot{z}^1 & = & f^1(z^1, z^2) & + \sum_{i=1}^m g_i^1(z^1, z^2)u^i \\ \dot{z}^2 & = & f^2(z^2) & + \sum_{i=1}^m g_i^2(z^2)u^i \\ y & = & h(z^2). & \end{array}$$

Proof: The representation of the vector fields follows directly from Lemma 10.1.3. The condition that the outputs are only a function of the z^2 coordinates results from the assumption that $\Delta \subset \ker\{dh^1, \ldots, dh^p\}$. This fact implies that the 1-forms $dh^1, \ldots, dh^p$ are contained in $\Delta^\perp$. In turn, this implies that

$$\frac{\partial h^i}{\partial z^j} = 0$$

for each i in the range $1 \leq i \leq p$ and each j in the range $1 \leq j \leq d$. Therefore, each output is only a function of the z^2 coordinates. □

Corollary 10.1.3 applies equally to the vector fields which define the standard extension of an affine control system to the state-time space, which can be written in matrix form as

$$\begin{bmatrix} g_1 & \cdots & g_m & f \\ 0 & & 0 & 1 \end{bmatrix}. \tag{10.5}$$

Lemma 10.1.4 also applies to the extended system 10.5 without change. Lemma 10.1.5 also applies, but with the understanding that the extended output function maps the state-time space to an output-time space. This extended mapping has the form $h_e : \mathcal{R}^n \times \mathcal{R} \rightarrow \mathcal{R}^p \times \mathcal{R} : (z, t) \rightarrow (h(z), t)$. As a consequence of this extension, we must modify Lemma 10.1.5 to require that the postulated distribution Δ satisfies the condition

$$\Delta \subset \ker\{dh^1, \ldots, dh^p, dt\}.$$

Lemma 10.1.4 assumes the existence of an involutive distribution which contains the input vector field $g_1, \ldots, g_m$ and is invariant with respect to the vector fields $g_1, \ldots, g_m$, and f. Such a distribution always exists since we can take Δ to be the distribution defined at each point $p \in M$ by $\Delta(p) = T_p M$. This distribution obviously satisfies the requirements of Lemma 10.1.4, but the resulting coordinates, which can be taken to be the same as the original coordinates, provide no new insight into the structure of the system. A more useful representation can be found by applying the closure properties of Lemma 10.1.2 to find the smallest distribution which satisfies the conditions of Lemma 10.1.4. To this end, let Δ be a fixed distribution and $f_1, \ldots, f_m$ be a collection of smooth vector fields which are defined over a manifold M. Let $\mathcal{S}^{[\Delta, TM]}(M)$ denote the collection of all smooth distributions containing Δ. We will to consider the subset of $\mathcal{S}^{[\Delta, T_p M]}(M)$ consisting of all smooth distributions which are involutive, contain Δ, and are invariant with respect to $f_1, \ldots, f_m$. The following lemma ensures that this set has a well-define minimal element.

Lemma 10.1.6 *The set* $\mathcal{S}_f^{[\Delta, TM]}(M)$ *defined by*

$$\mathcal{S}_f^{[\Delta, TM]}(M) := \{D \in \mathcal{S}^{[\Delta, TM]}(M) \mid [D, f_i] \subset D,\ [D, D] \subset D \ \ 1 \leq i \leq m\}$$

contains a unique involutive distribution $\hat{D} \in \mathcal{S}_f^{[\Delta, TM]}(M)$ *of minimal dimension.*

Proof: If Δ_1 and Δ_2 are any two involutive distributions, then their intersection is also involutive, and, from condition 3 in Lemma 10.1.2, we know that if Δ_1 and Δ_2 are both invariant with respect to f, then so is their intersection. This implies that the set $\mathcal{S}_f^{[\Delta, TM]}(M)$ contains a unique element which is contained in every other element of $\mathcal{S}_f^{[\Delta, TM]}(M)$. □

Lemma 10.1.5 assumes the the existence of a constant-dimensional, involutive distribution which is contained in the distribution $\ker\{dh^1, \ldots, dh^p\}$ and is invariant with respect to the vector fields $g_1, \ldots, g_m$, and f. Such a distribution always exists since we can take Δ to be the distribution defined at each point $p \in M$ by $\Delta(p) = \text{span}\{0\}$. Furthermore, the closure properties of Lemma 10.1.2 ensure that the set of all distribution which satisfy these conditions contains a unique maximal element.

In order to state this more formally, let Δ be a fixed distribution and $f_1, \ldots, f_m$ be a collection of smooth vector fields which are defined over a manifold M. Let $\mathcal{S}^{[0, \Delta]}(M)$ denote the set of all distributions which are contained in Δ. We are interested in the subset of $\mathcal{S}^{[0, \Delta]}(M)$ which consists of all distributions which are both contained in Δ and invariant with respect to the vector fields $f_1, \ldots, f_m$. The following lemma ensures that this set has a well-defined maximal element.

Lemma 10.1.7 *The set* $\mathcal{S}_f^{[0, \Delta]}(M)$ *defined by*

$$\mathcal{S}_f^{[0, \Delta]}(M) := \{D \in \mathcal{S}^{[0, \Delta]}(M) \mid [D, f_i] \subset D \ \ 1 \leq i \leq m\}$$

contains a unique distribution $\hat{D} \in \mathcal{S}_f^{[0, \Delta]}(M)$ *of maximal dimension. Furthermore, if the distribution* Δ *is involutive, then so is the distribution* $\hat{D}$.

Proof: The first statement follows immediately from the first condition in Lemma 10.1.2. If Δ is involutive, then

$$[f_i, \hat{D}] \subset \hat{D} \subset [\hat{D}, \hat{D}] \subset [\Delta, \Delta] \subset \Delta.$$

Since $\hat{D}$ is invariant with respect to the vector fields $f_1, \ldots, f_m$, the third condition in Lemma 10.1.2 ensures that $[\hat{D}, \hat{D}]$ is also invariant with respect to $f_1, \ldots, f_m$. Therefore, $[\hat{D}, \hat{D}] \in \mathcal{S}_f^{[0, \Delta]}(M)$. Finally, the maximality of $\hat{D}$ implies that $[\hat{D}, \hat{D}] \subset \hat{D}$. This proves the second statement. □

In Chapter 9, we discussed two different definitions of equivalence for affine control systems: equivalence with respect to state transformation and equivalence with respect to feedback transformation. If a distribution is invariant with respect to to the collection of vector fields $g_1, \ldots, g_m, f$

which define an affine control system, then this fact remains true no matter what local coordinates are used to describe the vector fields. Therefore, this property is invariant with respect to state transformations. However, this property is not preserved under the more general feedback transformations. Since feedback transformations are incorporated into the structure of the Grassmann bundle model an affine control system, it is necessary to consider a type of invariance which is preserved under feedback transformations. This topic is addressed in the next section.

10.2 Invariant Pairs of Distributions

Two distributions Δ and Γ are said to form an invariant pair if and only if

$$[\Delta, \Gamma] \subset \Delta + \Gamma. \tag{10.6}$$

If Δ and Γ form an invariant pair, then we will say that Δ is invariant with respect to Γ, and symmetrically that Γ is invariant with respect to Δ. Two codistributions I and J are said to form an invariant pair if and only if the distributions $I^{\perp}$ and $J^{\perp}$ form an invariant pair. If both Δ and Γ are involutive, then this property has the following geometric interpretation

Lemma 10.2.1 *Two involutive distributions Δ and Γ form an invariant pair if and only if the distribution $\Delta + \Gamma$ is involutive.*

Proof: If (Δ, Γ) is an invariant pair, then

$$[\Delta + \Gamma, \Delta + \Gamma] \quad = \quad [\Delta, \Delta] + [\Gamma, \Gamma] + [\Delta, \Gamma] \quad \subset \quad \Delta + \Gamma + [\Delta, \Gamma],$$

and using equation 10.6 we obtain

$$[\Delta + \Gamma, \Delta + \Gamma] \quad \subset \quad \Delta + \Gamma.$$

Conversely, if $\Delta + \Gamma$ is involutive, then since Δ and Γ are both subsets of the involutive distribution $\Delta + \Gamma$, we must have that

$$[\Delta, \Gamma] \subset [\Delta + \Gamma, \Delta + \Gamma] \subset \Delta + \Gamma.$$

□

Theorem 10.2.1 *Let Δ_1 and Δ_2 be two distributions. The following statements are equivalent.*

1. $[\Delta_1, \Delta_2] \subset \Delta_1 + \Delta_2$.
2. $L_{\Delta_1}(\Delta_1^{\perp} \cap \Delta_2^{\perp}) \subset \Delta_2^{\perp}$.
3. $L_{\Delta_2}(\Delta_1^{\perp} \cap \Delta_2^{\perp}) \subset \Delta_1^{\perp}$.

Proof: Using Lemma 7.2.4, we have that $(\Delta_1 + \Delta_2)^{\perp} = \Delta_1^{\perp} \cap \Delta_2^{\perp}$. Therefore, we find that

$$\begin{aligned} & [\Delta_1, \Delta_2] \subset \Delta_1 + \Delta_2 \\ \Leftrightarrow \quad & \Delta_1^{\perp} \cap \Delta_2^{\perp} \subset [\Delta_1, \Delta_2]^{\perp} \\ \Leftrightarrow \quad & [\Delta_1, \Delta_2] \lrcorner (\Delta_1^{\perp} \cap \Delta_2^{\perp}) = 0. \end{aligned}$$

Using Cartan's formula and the facts that $\Delta_1 \lrcorner (\Delta_1^{\perp} \cap \Delta_2^{\perp}) = 0$ and $\Delta_2 \lrcorner (\Delta_1^{\perp} \cap \Delta_2^{\perp}) = 0$, we also find that

$$[\Delta_1, \Delta_2] \lrcorner (\Delta_1^{\perp} \cap \Delta_2^{\perp}) = \Delta_2 \lrcorner \Delta_1 \lrcorner d((\Delta_1^{\perp} \cap \Delta_2^{\perp}).$$

Therefore,

$$\begin{aligned} & [\Delta_1, \Delta_2] \lrcorner (\Delta_1^{\perp} \cap \Delta_2^{\perp}) = 0 \quad \Leftrightarrow \quad \Delta_1 \lrcorner \Delta_2 \lrcorner d(\Delta_1^{\perp} \cap \Delta_2^{\perp}) = 0 \\ \Leftrightarrow \quad & \Delta_1 \lrcorner d(\Delta_1^{\perp} \cap \Delta_2^{\perp}) \subset \Delta_2^{\perp} \quad \Leftrightarrow \quad \Delta_2 \lrcorner d(\Delta_1^{\perp} \cap \Delta_2^{\perp}) \subset \Delta_1^{\perp}. \end{aligned}$$

Finally, using formula 7.8 and the facts that $\Delta_1 \lrcorner (\Delta_1^\perp \cap \Delta_2^\perp) = 0$ and $\Delta_2 \lrcorner (\Delta_1^\perp \cap \Delta_2^\perp) = 0$, it is easy to show that

$$\Delta_1 \lrcorner d((\Delta_1^\perp \cap \Delta_2^\perp) = L_{\Delta_1}(\Delta_1^\perp \cap \Delta_2^\perp) \text{ and } \Delta_2 \lrcorner d((\Delta_1^\perp \cap \Delta_2^\perp) = L_{\Delta_2}(\Delta_1^\perp \cap \Delta_2^\perp).$$

Hence, the three conditions are equivalent. □

The following lemma provides some useful closure properties of invariant pairs.

Lemma 10.2.2

1. *If (Δ_1, Γ) and (Δ_2, Γ) are two invariant pairs of distributions, then $(\Delta_1 + \Delta_2, \Gamma)$ is also an invariant pair.*
2. *If (Δ, Γ) is an invariant pair of distributions, then $([\Delta, \Delta], \Gamma)$ is also an invariant pair.*

Proof: To prove statement 1, we note that since (Δ_1, Γ) and (Δ_2, Γ) are both invariant pairs,

$$\begin{aligned} [\Delta_1, \Gamma] &\subset \Delta_1 + \Gamma \\ [\Delta_2, \Gamma] &\subset \Delta_2 + \Gamma \\ \Rightarrow [\Delta_1, \Gamma] + [\Delta_2, \Gamma] &\subset \Delta_1 + \Delta_2 + \Gamma \\ \Rightarrow [(\Delta_1 + \Delta_2), \Gamma] &\subset (\Delta_1 + \Delta_2) + \Gamma. \end{aligned}$$

Consequently, $(\Delta_1 + \Delta_2, \Gamma)$ also forms an invariant pair.

To prove statement 2, we note that if (Δ, Γ) is an invariant pair, then

$$\begin{aligned} [\Delta, \Gamma] &\subset \Delta + \Gamma \\ \Rightarrow [\Delta, [\Delta, \Gamma]] &\subset [\Delta, \Delta] + [\Delta, \Gamma] \\ \Rightarrow [\Delta, [\Delta, \Gamma]] &\subset [\Delta, \Delta] + \Gamma. \end{aligned}$$

Therefore, if we pick any two vectors $v_1, v_2 \in \Delta$ and any vector $w \in \Gamma$, they will satisfy the equations

$$\begin{aligned} [v_1, [v_2, w]] &\subset [\Delta, \Delta] + \Gamma \\ [v_2, [v_1, w]] &\subset [\Delta, \Delta] + \Gamma \\ \Rightarrow [v_1, [v_2, w]] - [v_2, [v_1, w]] &\subset [\Delta, \Delta] + \Gamma. \end{aligned} \tag{10.7}$$

The vectors v_1, v_2 and r are also related through the Jacobi identity

$$[v_1, [v_2, w]] - [v_2, [v_1, w]] = [[v_1, v_2], w].$$

Substituting the right hand side of this expression into equation 10.7 gives

$$[[v_1, v_2], w] \subset [\Delta, \Delta] + \Gamma.$$

Since this equation holds for arbitrary $v_1, v_2 \in \Delta$ and arbitrary $w \in \Gamma$, it implies that

$$[[\Delta, \Delta], \Gamma] \subset [\Delta, \Delta] + \Gamma.$$

Therefore, $([\Delta, \Delta], \Gamma)$ is also an invariant pair. □

In the previous section, we discussed the invariance of distributions and codistributions with respect to vector fields. We are now ready to show how this material is related to the concept of an invariant pair of distributions.

Assume that we are given a pair of codistributions (I, J) which form an invariant pair. Then

$$L_{J^\perp}(I \cap J) \subset I.$$

If we make the additional assumption that J is involutive, then we have that

$$L_{J^\perp}(I \cap J) \subset L_{J^\perp} J = C(J) = J,$$

so we find that

$$L_{J^\perp}(I \cap J) \subset (I \cap J).$$

This implies that $J^\perp \subset A(I \cap J)$ or equivalently that $C(I \cap J) \subset J$. If the retracting space has constant dimension, then Theorem 8.3.2 says that there exists a local coordinate chart and a set of generators for $(I \cap J)$ whose coefficients are only functions of the coordinates whose differentials span $C(I \cap J)$.

Let I have dimension r, J have dimension p, and $I \cap J$ have dimension q. Let $\{dz^1, \ldots, dz^p\}$ be a collection of one-forms which span J and let $\{dx^1, \ldots, dx^{(n-p)}\}$ be a collection of one-forms which, taken together with the dz^i, form a basis of T^*M. Any 1-form ω can be expressed with respect to this cobasis as

$$\omega^i = \sum_{j=1}^{p} a^i_{1j} dz^j + \sum_{k=1}^{(n-p)} a^i_{2k} dx^k.$$

By Theorem 8.3.2 and the involutivity of J, there exists a basis of one-forms $\theta^1, \ldots, \theta^q$ which span $I \cap J$ and which can be expressed solely in terms of the z^i coordinates

$$\theta^i = \sum_{j=1}^{p} b^i_j(z) dz^j.$$

Suppose that we select an additional set of one-forms $\{\omega^1, \ldots, \omega^{r-q}\}$ which, taken together with the θ^i, form a cobasis of I and an additional subset of the dx and dz so that the entire collection forms a cobasis of T^*M. This cobasis can be represented as the matrix of coefficient functions

$$\begin{bmatrix} A_1(x,z) & A_2(x,z) & A_3(x,z) & A_4(x,z) \\ 0 & I & 0 & 0 \\ 0 & 0 & B_1(z) & B_2(z) \\ 0 & 0 & 0 & I \end{bmatrix}$$

expressed with respect to the cobasis

$$\{dx^1, \ldots, dx^{n-p}, dz^1, \ldots, dz^p\}$$

By inverting this matrix, we obtain a matrix of coefficient functions which represents a basis of TM

$$\begin{bmatrix} A_1^{-1} & -A_1^{-1}A_2 & (-A_1^{-1}A_3B_1^{-1}) & (A_1^{-1}A_3B_1^{-1}B_2 - A_1^{-1}A_4) \\ 0 & I & 0 & 0 \\ 0 & 0 & B_1^{-1} & -B_1^{-1}B_2 \\ 0 & 0 & 0 & I \end{bmatrix}$$

expressed with respect to the standard coordinate basis

$$\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{n-p}}, \frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^p} \right\}.$$

Consequently, $I^\perp$ is spanned by the columns of the matrix

$$\begin{bmatrix} -A_1^{-1}A_2(x,z) & (A_1^{-1}A_3B_1^{-1}B_2 - A_1^{-1}A_4)(x,z) \\ I & 0 \\ 0 & -B_1^{-1}B_2(z) \\ 0 & I \end{bmatrix} \tag{10.8}$$

which is also defined with respect to the standard coordinate basis.

To recap, we began the previous construction with two codistributions I and J. We supposed that these distributions form an invariant pair, that the codistribution J is involutive, and that the distribution $C(I \cap J)$ has constant dimension. Based on this information, we deduced the existence of a coordinate chart (x, z) such that the differentials $dz^1, \ldots, dz^p$ span $C(I \cap J)$ and the existence of set of vector fields which span the distribution $I^{\perp}$ and have the form of equation 10.8. Expressed in the (x, z) coordinates, these vector fields have a triangular decomposition similar to that of Corollary 10.1.3. The following lemma makes this statement more precise.

Lemma 10.2.3 *The codistribution J is invariant with respect to the set of vectors which make up the frame 10.8.*

Proof: Let f_i denote the ith column of the matrix 10.8. Any 1-form $\alpha \in J$ can be expressed as

$$\alpha = \sum_{i=1}^{q} c_i(x, z)\theta^i + \sum_{j=1}^{p-q} h_j(x, z)dz^i.$$

We need to verify that α satisfies

$$L_{f_i}\alpha \in J.$$

Writing this out in coordinates, we find that

$$\begin{aligned} L_{f_i}\alpha &= \left(\sum_{i=1}^{q}(f_i \lrcorner\, dc_i) \wedge \theta^i + \sum_{i=1}^{q} c_i(x, z)(f_i \lrcorner\, d\theta^i) + \sum_{j=1}^{p-q}(f_i \lrcorner\, dh_j) \wedge dz^i\right) \\ &- \left(\sum_{i=1}^{q} dc_i \wedge (f_i \lrcorner\, \theta^i) + \sum_{j=1}^{p-q} dh_j \wedge (f_i \lrcorner\, dz^i)\right) \\ &+ d\left(\sum_{i=1}^{q} c_i(x, z)(f_i \lrcorner\, \theta^i) + \sum_{j=1}^{p-q} h_j(x, z)(f_i \lrcorner\, dz^i)\right). \end{aligned}$$

If $1 \leq i \leq r - q$, then $f_i \in J^{\perp}$, and this implies that the last two lines of this summation are identically zero. The first and third terms on the first line lie in the span of J, and, due to the special structure of the θ^i, so does the second term on the first line. If $r - q + 1 \leq i \leq r$ then by construction all the terms $f_i \lrcorner\, \theta^j$ and $f_i \lrcorner\, dz^j$ are constants, and the last two lines cancel each other. Again, all the terms on the first line lie in the span of J, so the proof is complete. □

This result also has a nice geometric interpretation. If J is an involutive, constant dimensional distribution, then the functions $z^1, \ldots, z^p$ locally define a smooth surjection $\pi : U \subset M \to \mathcal{R}^p$. Which is defined with respect to the x, z coordinate system by $\pi : (x, z) \to z$. Since the 1-forms θ^i depend only on the z coordinates, they are each equal to the pull-back of some form $\hat{\theta}^i$ defined on $\mathcal{R}^p$.

$$\theta^i = \pi^* \hat{\theta}^i.$$

Furthermore, there is a bijective correspondence between the set of all 1-dimensional integral elements over a point $z \in \mathcal{R}^p$ of the exterior differential system generated by $\{\hat{\theta}^1, \ldots, \hat{\theta}^q\}$ on $\mathcal{R}^p$, and the $n - p + 1$-dimensional integral elements of the exterior differential system generated by $\{\theta^1, \ldots, \theta^q\}$ on M over any point $p \in \pi^{-1}(z)$.

The concept of an invariant pair is also related to the concept of a controlled invariant distribution. We begin by recalling the definition of a controlled invariant distribution in the standard theory.

Definition 10.2.1 *A distribution Δ is said to be controlled invariant with respect to an affine control system described by the vector fields $f, g_1, \ldots, g_m$ defined over a smooth manifold M if and only if there exists an affine state feedback $u^i = \alpha^i(x) + \sum_{j=1}^m \beta^i_j(x) v^j$ defined such that the transformed system*

$$\begin{aligned} \hat{f} &= f + \sum_{j=1}^m g_j \alpha^j \\ \hat{g}_i &= \sum_{j=1}^m g_j \beta^j_i \end{aligned}$$

satisfies

$$\begin{aligned} \left[\hat{f}, \Delta\right] &\subset \Delta \\ [\hat{g}_1, \Delta] &\subset \Delta \\ &\vdots \\ [\hat{g}_m, \Delta] &\subset \Delta. \end{aligned}$$

There is a geometric test to see if a given distribution is controlled invariant which does not depend explicitly on the feedback transformation.

Lemma 10.2.4 (Isidori) *A necessary condition for a distribution Δ to be controlled invariant is that*

$$\begin{aligned} [f, \Delta] &\subset \Delta + G \\ [g_1, \Delta] &\subset \Delta + G \\ &\vdots \\ [g_m, \Delta] &\subset \Delta + G \end{aligned}$$

where G is the distribution defined by

$$G := span\{g_1, \ldots, g_m\}.$$

If the distribution Δ is involutive, and the distributions Δ, G, and $\Delta \cap G$ are all nonsingular on an open subset $U \subset M$, then this condition is also sufficient for Δ to be controlled invariant.

Proof: See Isidori [4] pages 311-319. □

The following lemma establishes the connection between the concept of controlled invariance and the concept of an invariant pair of distributions.

Lemma 10.2.5 *Let $f, g_1, \ldots, g_m$ be a collection of vector fields which defined an affine control system over a smooth manifold M. Let F be a the smooth distribution defined over $M \times \mathcal{R}$ by*

$$F := span \begin{bmatrix} g_1 & \cdots & g_m & f \\ 0 & & 0 & 1 \end{bmatrix}.$$

Let Δ be a smooth distribution defined over M, and let Δ_e be the smooth distribution defined over $M \times \mathcal{R}$ by

$$\Delta_e = \left(\pi^*(\Delta^\perp) + span\{dt\}\right)^\perp$$

where π is the canonical projection $\pi : M \times \mathcal{R} \rightarrow M$. Then the distributions (F, Δ_e) form an invariant pair if and only if

$$\begin{aligned} [f, \Delta] &\subset \Delta + G \\ [g_1, \Delta] &\subset \Delta + G \\ &\vdots \\ [g_m, \Delta] &\subset \Delta + G. \end{aligned} \tag{10.9}$$

Proof: Any vector field $w(x,t) \in F$ can be written in the form

$$w(x,t) = \alpha(x,t) \begin{bmatrix} f(x) \\ 1 \end{bmatrix} + \sum_{j=1}^{m} \alpha(x,t) \begin{bmatrix} g_j(x) u^j(x,t) \\ 0 \end{bmatrix} \tag{10.10}$$

and any vector field $v(x,t) \in \Delta_e$ can be written in the form

$$v(x,t) = \begin{bmatrix} v(x,t) \\ 0 \end{bmatrix},$$

so the Lie bracket of any $w(x,t) \in F$ with any $v(x,t) \in \Delta_e$ can be written out as

$$[w(x,t), v(x,t)] = \alpha(x,t) \begin{bmatrix} \frac{\partial v}{\partial x} & \frac{\partial v}{\partial t} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f(x) + g_i(x) u^i(x,t) \\ 1 \end{bmatrix} \tag{10.11}$$

$$-\alpha(x,t) \begin{bmatrix} \frac{\partial f}{\partial x} + \frac{\partial g_i}{\partial x} u^i + g_i(x) \frac{\partial u^i}{\partial x} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v(x,t) \\ 0 \end{bmatrix} - L_v \alpha(x,t) \begin{bmatrix} f(x) + g_i(x) u^i \\ 1 \end{bmatrix}.$$

which can be rewritten as

$$[w, v] = \alpha \begin{bmatrix} \frac{\partial v}{\partial x} f(x) - \frac{\partial f}{\partial x} v(x,t) + \frac{\partial v}{\partial x} g_i(x) u^i(x,t) - \frac{\partial g_i}{\partial x} u^i(x,t) v(x,t) \\ 0 \end{bmatrix}$$

$$- \alpha \begin{bmatrix} g_i(x) \frac{\partial u^i}{\partial x} v(x,t) \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} \frac{\partial v}{\partial t} \\ 0 \end{bmatrix} - L_v \alpha \begin{bmatrix} f(x) + g_i(x) u^i(x,t) \\ 1 \end{bmatrix}. \tag{10.12}$$

If (F, Δ_e) forms an invariant pair, then

$$[F, \Delta_e] \subset F + \Delta_e. \tag{10.13}$$

Consequently,

$$[w, v] \in \operatorname{span} \left\{ \begin{bmatrix} g_1 & \cdots & g_m & f \\ 0 & & 0 & 1 \end{bmatrix} + \begin{bmatrix} \Delta \\ 0 \end{bmatrix} \right\}. \tag{10.14}$$

Comparing equations 10.14 and 10.12, it is apparent that if we fix a time $t = t_0$, and restrict the vector fields to the submanifold $M \times \{t_0\}$, then the condition 10.13 implies the conditions 10.9. Conversely, if 10.9 is assumed to hold, then condition 10.13 must also hold. Note that in equation 10.12, the vector

$$v_t := \begin{bmatrix} \frac{\partial v}{\partial t} \\ 0 \end{bmatrix}$$

must always satisfy $\pi(v_t) \in \Delta$ since the distribution Δ_e is not dependent on time. □

If the distribution Δ on M is involutive, then so is the distribution Δ_e on $M \times \mathcal{R}$. In this case, we get a nice geometric interpretation of the notion of a controlled invariant distribution. At a point $(x,t) \in M \times \mathcal{R}$, suppose that $\Delta_e(p,t)$ is a k-dimensional subspace of $T_{(p,t)}(M \times \mathcal{R})$. Consider the collection $G_{k+1}^{[\Delta_e, (\Delta_e + F)]}(M \times \mathcal{R})|_{(p,t)}$ of all $k+1$-dimensional subspaces of $T_{(p,t)}(M \times \mathcal{R})$ which contain $\Delta_e|_(p,t)$ and are contained in $(\Delta_e + F)|_{(p,t)}$. We have the following result.

Lemma 10.2.6 *There exists an output space $\mathcal{R}^p \times \mathcal{R}$ and an output mapping $\lambda : M \times \mathcal{R} \rightarrow \mathcal{R}^p \times \mathcal{R}$ $(x,t) \rightarrow (\lambda(x), t)$ such that*

1. *$\Delta_e = ker(d\lambda) = span\{d\lambda^1, \ldots, d\lambda^p, dt\}$.*
2. *The system $\Delta_e^\perp \cap F^\perp$ is spanned by a set of 1-forms which are the pull-backs of a set of 1-forms Ω defined on $\mathcal{R}^p \times \mathcal{R}$.*
3. *There exists a bijection between the set $G_{k+1}^{[\Delta_e,(\Delta_e+F)]}(M\times\mathcal{R})|_{(p,t)}$ and the set $G_1^{[0,\lambda_*(\Delta_e+F)]}(\mathcal{R}^p \times \mathcal{R})|_{(\lambda(x),t)}$ consisting of all one-dimensional subspaces satisfying $\Delta_e^\perp \cap F^\perp$ at at the point $(\lambda(x),t) \in \mathcal{R}^p \times \mathcal{R}$.*

Proof: The first statement is an immediate consequence of the involutivity of Δ_e.

To prove the second statement, we will use that fact that (F, Δ_e) is an invariant pair and the fact that Δ_e is involutive. Since (F, Δ_e) is an invariant pair,

$$L_{\Delta_e}(\Delta_e^\perp \cap F^\perp) \subset F^\perp,$$

and since Δ_e is involutive,

$$L_{\Delta_e}(\Delta_e^\perp \cap F^\perp) \subset L_{\Delta_e}\Delta_e^\perp \subset \Delta_e.$$

Taking the intersection of the last two equations, we find that

$$L_{\Delta_e}(\Delta_e^\perp \cap F^\perp) \subset (\Delta_e^\perp \cap F^\perp).$$

Therefore,

$$\Delta_e \subset A(\Delta_e^\perp \cap F^\perp) \Leftrightarrow C(\Delta_e^\perp \cap F^\perp) \subset \Delta_e^\perp.$$

Using Theorem 8.3.2 we can conclude that $(\Delta_e^\perp \cap F^\perp)$ is spanned by a set of 1-forms whose coefficients only involve the coordinates $\lambda^1, \ldots, \lambda^p$, and t. Consequently, each of these forms represents the pull-back of a form on $\mathcal{R}^p \times \mathcal{R}$. We will use Ω to denote the codistribution spanned by these forms on $\mathcal{R}^p \times \mathcal{R}$.

Since $(\Delta_e^\perp \cap F^\perp) = \lambda^*(\Omega)$, we must also have that $\lambda_*(\Delta_e + F) = \Omega^\perp$. Furthermore, by construction $\Delta_e^\perp = \ker(\lambda_*)$, so any subspace $D \in G_{k+1}^{[\Delta_e,(\Delta_e+F)]}(M \times \mathcal{R})|_{(p,t)}$ must map to a one-dimensional subspace $\lambda_*(D) \in G_1^{[0,\lambda_*(\Delta_e+F)]}(\mathcal{R}^p \times \mathcal{R})|_{(\lambda(x),t)}$. Since this is an onto mapping, Lemma 5.4.1 implies that these two Grassmann manifolds are isomorphic. □

Let Δ and Γ be two distribution which are defined over a manifold M. Let $\mathcal{S}^{[0,\Delta]}(M)$ denote the set of all distributions which are contained in Δ. We are interested in the subset of $\mathcal{S}^{[0,\Delta]}(M)$ which consists of all the distributions which are contained in Δ and invariant with respect to Γ.

Lemma 10.2.7 *The set $\mathcal{S}_\Gamma^{[0,\Delta]}(M)$ defined by*

$$\mathcal{S}_\Gamma^{[0,\Delta]}(M) := \{D \in \mathcal{S}^{[0,\Delta]}(M) \mid [D, \Gamma] \subset D + \Gamma\}$$

contains a unique distribution $\hat{D} \in \mathcal{S}_\Gamma^{[0,\Delta]}(M)$ of maximal dimension. Furthermore, if the distribution Δ is involutive, then so is the distribution $\hat{D}$.

Proof: The first statement follows immediately from equation 1 of Lemma 10.2.2. If Δ is involutive, then

$$[\hat{D}, \hat{D}] \subset [\Delta, \Delta] \subset \Delta.$$

Since $\hat{D}$ is invariant with respect to Γ, equation 2 of Lemma 10.2.2 ensures that $[\hat{D}, \hat{D}]$ is also invariant with respect to Γ. Therefore, $[\hat{D}, \hat{D}] \in \mathcal{S}_\Gamma^{[0,\Delta]}(M)$. Finally, the maximality of $\hat{D}$ implies that $[\hat{D}, \hat{D}] \subset \hat{D}$. This proves the second statement. □

10.2.1 The Filtration Associated With The Largest Distribution Contained in Δ

Given an arbitrary distribution Δ, Lemma 10.2.7 says that there exists a unique distribution $\hat{D}$ which is contained in Δ, invariant with respect to Γ, and maximal in the sense that any other distribution which satisfies the first two properties is contained in $\hat{D}$. In what follows we will develop an algorithm which can be used to explicitly compute $\hat{D}$.

We begin by setting $J_0 = \Delta^\perp$ and recursively define the codistributions

$$J_{k+1} = L_\Gamma(J_k \cap \Gamma^\perp) + J_k. \tag{10.15}$$

If there exists a integer k^* such that $J_{k^*-1} \neq J_{k^*}$ and $J_{k^*} = J_{k^*+1}$, then

$$L_\Gamma(J_{k^*} \cap \Gamma^\perp) \subset J_{k^*},$$

and the codistributions

$$J_0 \subset J_1 \subset \ldots \subset J_{k_*}$$

form a filtration of finite length.

If each of the codistributions $J_k \cap \Gamma^\perp$ and J_k has constant dimension on some neighborhood of a point $p \in M$, then p is called a regular point of Algorithm 10.15.

Theorem 10.2.2 *If the distribution Δ is involutive, and a point $p \in M$ is a regular point of Algorithm 10.15, then each of the codistributions in the filtration*

$$J_0 \subset J_1 \subset \cdots \subset J_{k_*}$$

is involutive.

Proof: Since we are assuming that Δ is involutive, the codistribution J_0 is obviously involutive. Assume that the codistribution J_k is involutive. Since J_k and $J_k \cap \Gamma^\perp$ are constant dimensional, locally there exist smooth functions $z^1, \ldots, z^{p_k}$ such that

$$J_k = \operatorname{span}\{dz^1, \ldots, dz^{p_k}\}$$

and smooth one-forms θ^i $1 \leq i \leq r_k$ which satisfy

$$J_k \cap \Gamma^\perp = \operatorname{span}\{\theta^1, \ldots, \theta^{r_k}\}.$$

If $\Gamma^\perp$ is m-dimensional, then we can select an additional set of one-forms $\omega^1, \ldots, \omega^{m-r_k}$ so that

$$\Gamma^\perp = \operatorname{span}\{\omega^1, \ldots, \omega^{m-r_k}, \theta^1, \ldots, \theta^{r_k}\}.$$

If we also select an additional set of smooth functions $x^1, \ldots, x^{n-p_k}$ then, with respect to the basis

$$\{dx^1, \ldots, dx^{n-p_k}, dz^1, \ldots, dz^{p_k}\}, \tag{10.16}$$

each ω^i can be expressed as

$$\omega^i = \sum_{j=1}^{n-p_k} c^i_j(x,z)dx^j + \sum_{j=1}^{p_k} b^i_j(x,z)dz^j,$$

and each θ^i can be expressed as

$$\theta^i = \sum_{j=1}^{p} c^i_j(x,z)dz^j.$$

We can select a subset of the dx's and dz's so that, taken together with the θ's and ω's, we have a basis for T^*M. The coefficients of this basis form a matrix of smooth functions

$$\begin{bmatrix} A_1(x,z) & A_2(x,z) & B_1(x,z) & B_2(x,z) \\ 0 & I & 0 & 0 \\ 0 & 0 & C_1(x,z) & C_2(x,z) \\ 0 & 0 & 0 & I \end{bmatrix}$$

which is expressed with respect to the cobasis

$$\{dx^1, \ldots, dx^{n-p_k}, dz^1, \ldots, dz^{p_k}\}.$$

Since A_1 and C_1 are nonsingular, we can always produce an equivalent basis of the form

$$\begin{bmatrix} I & A_2(x,z) & B_1(x,z) & B_2(x,z) \\ 0 & I & 0 & 0 \\ 0 & 0 & I & C_2(x,z) \\ 0 & 0 & 0 & I \end{bmatrix}.$$

If we invert this matrix, we obtain a matrix of coefficient functions

$$\begin{bmatrix} I & -A_2 & -B_1 & (B_1C_2 - B_2) \\ 0 & I & 0 & 0 \\ 0 & 0 & I & -C_2 \\ 0 & 0 & 0 & I \end{bmatrix}$$

which represents a basis of TM expressed with respect to the standard coordinate basis

$$\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{n-p_k}}, \frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^{p_k}} \right\}.$$

Consequently, Γ is spanned by the columns of the matrix

$$\begin{bmatrix} -A_2(x,z) & (B_1C_2 - B_2)(x,z) \\ I & 0 \\ 0 & -C_2(x,z) \\ 0 & I \end{bmatrix}. \tag{10.17}$$

In order to compute J_{k+1} we must compute $L_f\alpha$ for every $f \in \Gamma$ and every $\alpha \in J_k \cap \Gamma^\perp$. An such α can be written as

$$\alpha = \sum_{i=1}^{q} c_i(x,z)\theta^i.$$

Since $f \lrcorner \alpha = 0$, $L_f\alpha = f \lrcorner d\alpha$, this operation can be viewed as a mapping taking $T_pM \times T_p^*M \to T_p^*M$ which is linear with respect to the ring $C^\infty(M)$ in the first argument. Therefore, it suffices to work with any basis of Γ, and we will choose the basis described by the matrix 10.17. We will denote the ith column of this matrix by f_i and compute $L_{f_i}\alpha$. Writing this out in coordinates, we find that

$$L_{f_i}\alpha \quad = \quad \left(\sum_{i=1}^{q_k} (f_i \lrcorner dc_i) \wedge \theta^i + \sum_{i=1}^{q_k} c_i(x,z)(f_i \lrcorner d\theta^i) \right)$$

The first term on the right hand side of this equation is in the span of J_k, so that any 1-form in J_{k+1} can be expressed as a linear combination of the 1-forms $f_i \lrcorner d\theta^j$ mod J_k. Therefore, we only need to consider the one-forms $L_{f_i}\theta^j$. Each θ^j has the form

$$\theta^j \quad = \quad dz^j + \sum_{k=1}^{p_k - r_k} c_k^j(x,z)dz^k,$$

so the Lie derivative can be written out as

$$\begin{aligned} L_{f_i}\theta^j &= \left(\sum_{i=1}^{p_k - r_k} (f_i \lrcorner dc_k^j) \wedge dz^k \right) \\ &- \left(\sum_{i=1}^{p_k - r_k} dc_k^j \wedge (f_i \lrcorner dz^k) \right) \\ &+ d(f_i \lrcorner dz^j) \\ &+ d\left(\sum_{i=1}^{p_k - r_k} c_k^j (f_i \lrcorner dz^k) \right). \end{aligned} \tag{10.18}$$

By construction, each of the terms $f_i \lrcorner dz^k$ has a constant value of either 1 or 0, so the second and fourth lines cancel. Each term in the first line lies in the span of J_k, so each of these expressions can be simplied to

$$L_{f_i}\theta^j = d(f_i \lrcorner dz^j) \bmod J_k. \tag{10.19}$$

Thus, the codistribution J_{k+1} is spanned by the differentials

$$J_{k+1} = \text{span}\{d(f_i \lrcorner dz^j)\} + J_k.$$

Since J_{k+1} is locally spanned by exact differentials, it is also involutive. □

Since $f_i \lrcorner \theta^j \equiv 0$, we have that

$$f_i \lrcorner dz^j + \sum_{i=1}^{p_k - r_k} c_k^j (f_i \lrcorner dz^k) \equiv 0.$$

By construction, we also have that $f_i \lrcorner dz^k = \delta_i^k$. Therefore, we find that

$$L_{f_i} z^j = f_i \lrcorner dz^j = c_i^j.$$

Corollary 10.2.1 *If the codistribution J is involutive, and the point x is a regular point of the controlled invariant distribution algorithm, then*

$$C(J_k \cap F^\perp) \subset J_{k+1}.$$

Proof: Since $J_{k+1} = J_k + L_F(F^\perp \cap J_k)$, we must have that $L_F(F^\perp \cap J_k) \subset J_{k+1}$. From Theorem 10.2.2 we know that J_k is involutive, so

$$L_{J_k^\perp}(F^\perp \cap J_k) \subset L_{J_k^\perp} J_k \subset J_k \subset J_{k+1}.$$

Forming the subspace sum of the last two equations, we find that

$$L_{(F+J_k^\perp)}(F^\perp \cap J_k) = C(F^\perp \cap J_k) \subset J_{k+1}.$$

□

Corollary 10.2.2 *There exists a coordinate system on M and integers $r_0, \ldots, r_{k_*}$ and $s_0, \ldots, s_{k_*}$ such that*

1. $s_i \leq r_i$, *for* $1 \leq i \leq k_*$.
2. *span*$\{dx^1, \ldots, dx^{r_k}\} = J_k$.
3. *span*$\{\omega^1, \ldots, \omega^{s_k}\} = F^\perp \cap J_k$.
4. $\omega_i = \sum_{j=1}^{r_k} a_j^i(x^1, \ldots, x^{r(k+1)}) dx^j$.

Proof: The proof follows immediately from Corollary 10.2.1 and Theorem 8.3.2. □

10.2.2 The Largest Controlled Invariant Distribution Contained in ker(dh)

In the vector field theory, an invariant codistribution of particular importance is the largest controlled invariant distribution contained in $ker(dh)$. Isidori presents an algorithm for computing this distribution which proceeds as follows

1. Set $\Omega_0 = \text{span}\{dh^1, \dots dh^p\}$.
2. Set $\Omega_{k+1} = \Omega_k + L_f(\Omega_k \cap G^\perp) + \sum_{i=1}^m L_{g_i}(\Omega_k \cap G^\perp)$.

Isidori also presents the following lemma

Lemma 10.2.8 (Isidori) *Suppose there exists an integer k^* such that $\Omega_{k^*+1} = \Omega_{k^*}$, if Ω_{k^*} and $\Omega_{k^*} \cap G^\perp$ are smooth, then $\Omega_{k^*}^\perp$ is the maximal controlled invariant distribution contained in $ker(dh)$.*

This algorithm closely resembles algorithm 10.15. This relationship is made precise by the following lemma.

Lemma 10.2.9 *Let F denote distribution spanned by the columns of the matrix*

$$F := span \begin{bmatrix} g_1 & \cdots & g_m & f \\ 0 & & 0 & 1 \end{bmatrix}.$$

Suppose that the codistribution $J_0 = ker(dh_e) = span\{dh_1, \dots, dh_p, dt\}$ is defined over $M \times \mathcal{R}$. If there exists a k^* such that the algorithm 2 stabilizes, then Algorithm 10.15 which starts with the codistribution J_0 stabilizes, and each $J_k = \pi^*(\Omega_k) + span\{dt\}$.

Proof: By construction, the statement is true for $k = 0$. Suppose that it is true for some $k < k_*$, so that

$$\Omega_k = \text{span}\{\phi^1, \dots, \phi^{p_k}\}$$

and

$$J_k = \text{span}\{\phi^1, \dots, \phi^{p_k}, dt\}.$$

Assume that on M, the perp of the distribution G is spanned by the covectors

$$G^\perp = \text{span}\{\omega^1, \dots, \omega^{(n-m)}\},$$

and that the codistribution $\Omega_k \cap G^\perp$ is spanned by the covectors

$$\Omega_k \cap G^\perp = \text{span}\{\alpha^1, \dots, \alpha^{l_k}\}.$$

On $M \times \mathcal{R}$, we will have

$$J_k = \text{span}\{\phi^1, \dots, \phi^{p_k}, dt\},$$

$$F^\perp = \text{span}\{\omega^1 - (f \lrcorner \omega^1)dt, \dots, \omega^{p_k} - (f \lrcorner \omega^{p_k})dt\},$$

and

$$(F^\perp \cap J_k) = \text{span}\{\alpha^1 - (f \lrcorner \alpha^1)dt, \dots, \alpha^{l_k} - (f \lrcorner \alpha^{l_k})dt\}.$$

Using the controlled invariant distribution algorithm, we find that

$$\Omega_{k+1} = \text{span}\{\phi^1, \dots, \phi^{p_k}, L_f\alpha^1, \dots, L_f\alpha^{l_k}, L_{g_1}\alpha^1, \dots, L_{g_1}\alpha^{l_k}, \dots, L_{g_m}\alpha^1, \dots, L_{g_m}\alpha^{l_k}\}.$$

On $M \times \mathcal{R}$ let

$$g_1e := \begin{bmatrix} g_1 \\ 0 \end{bmatrix}, \cdots, g_me := \begin{bmatrix} g_m \\ 0 \end{bmatrix}, f_e := \begin{bmatrix} f \\ 1 \end{bmatrix}.$$

Clearly, these vector fields span F, so we can compute $L_F(F^\perp \cap J_k)$ by computing the covectors

$$\begin{aligned} L_{g_je}(\alpha^i - (f \lrcorner \alpha^i)dt &= L_{g_j}\alpha^i - L_{g_j}(f \lrcorner \alpha^i)dt \\ L_{f_e}(\alpha^i - (f \lrcorner \alpha^i)dt &= L_f\alpha^i - L_f(f \lrcorner \alpha^i)dt. \end{aligned}$$

Using these covectors, we find that

$$J_{k+1} = \{\phi^1, \dots, \phi^{p_k}, dt\} + L_F(F^\perp \cap J_k = \pi^*(\Omega_k) + \text{span}\{dt\}.$$

□

10.2.3 Controllability Distributions

This section introduces a special class of controlled invariant distributions which are called controllability distributions. Roughly speaking, these distributions characterize the submanifolds which can be controlled by a subset of the inputs. We will make use of this theory in Chapter 11 when we look at the noninteracting controls problem.

Lemma 10.2.10 *Given any two distribution Δ and Γ, there exists a unique distribution $\hat{D} \subset \Gamma$ of minimal dimension which satisfies*

1. $\Delta \cap \Gamma \subset \hat{D} \subset \Gamma$.
2. $[\Delta, \hat{D}] \cap \Gamma = \hat{D}$.

Proof: Let $\mathcal{T}(\Delta; \Gamma)$ denote the set of all distributions D satisfying

$$[\Delta, D] \cap \Gamma \subset D.$$

This set is not empty since

$$[\Delta, \Gamma] \cap \Gamma \subset \Gamma.$$

If $D_1, D_2 \in \mathcal{T}(\Delta; \Gamma)$, then

$$\Delta \cap \Gamma \subset D_1, D_2 \Rightarrow \Delta \cap \Gamma \subset D_1 \cap D_2$$

$$\begin{aligned} [\Delta, D_1] \cap \Gamma &\subset D_1 \\ [\Delta, D_2] \cap \Gamma &\subset D_2. \\ \Rightarrow [\Delta, (D_1 \cap D_2)] \cap \Gamma &\subset D_1 \cap D_2 \end{aligned}$$

Since $\mathcal{T}(\Delta; \Gamma)$ is nonempty and closed under smooth intersections, it must contain a unique minimal element.

Up to this point, we have shown that there exists a unique $\hat{D}$ of minimal dimension which satisfies $[\Delta, \hat{D}] \cap \Gamma \subset \hat{D}$. We still need to show equality. To do this, we begin by defining the distribution

$$K := [\Delta, \hat{D}] \cap \Gamma.$$

This distribution is a subset of $\hat{D}$, so we must have that

$$[\Delta, K] \cap \Gamma \subset [\Delta, \hat{D}] \cap \Gamma = K.$$

This distribution also contains $\Delta \cap \Gamma$, so it is a member of $K \in \mathcal{T}(\Delta; \Gamma)$. From the minimality of $\hat{D}$, we conclude that $K = [\Delta, \hat{D}] \cap \Gamma = \hat{D}$. □

Lemma 10.2.11 *If the distribution Γ is involutive, then so is the distribution $\hat{D}$.*

Proof: To prove this lemma, we will use the fact that, given any distribution D, the Cauchy characteristic of its perp, $A(D^\perp)$, is a subset of D. We will show that if Γ is involutive, then the condition

$$[\Delta, \hat{D}] \cap \Gamma = \hat{D}$$

implies that

$$\left[\Delta, A(\hat{D}^\perp)\right] \cap \Gamma \subset A(\hat{D}^\perp). \tag{10.20}$$

Equation 10.20 is equivalent to showing that the retracting space of $\hat{D}^\perp$ satisfies

$$(\left[\Delta, A(\hat{D}^\perp)\right] \cap \Gamma) \lrcorner C(\hat{D}^\perp) \equiv 0.$$

We compute

$$(\left[\Delta, A(\hat{D}^\perp)\right] \cap \Gamma) \lrcorner C(\hat{D}^\perp) =$$

$$(\left[\Delta, A(\hat{D}^\perp)\right] \cap \Gamma) \lrcorner L_{\hat{D}}(\hat{D}^\perp) =$$

$$(\left[\Delta, A(\hat{D}^\perp)\right] \cap \Gamma) \lrcorner \hat{D} \lrcorner d(\hat{D}^\perp) =$$

$$[[\Delta, A(\hat{D}^\perp)] \cap \Gamma, \hat{D}] \lrcorner \hat{D}^\perp \subset [[\Delta, A(\hat{D}^\perp)], \hat{D}] \cap \Gamma \lrcorner \hat{D}^\perp.$$

The last line follows from the involutivity of Γ. Using the Jacobi identity, we also get

$$[[\Delta, A(\hat{D}^\perp)], \hat{D}] \cap \Gamma \lrcorner \hat{D}^\perp = [[A(\hat{D}^\perp), \hat{D}], \Delta] \cap \Gamma \lrcorner \hat{D}^\perp.$$

Finally, using the fact that $[A(\hat{D}^\perp), \hat{D}] = \hat{D}$, we arrive at

$$[[A(\hat{D}^\perp), \hat{D}], \Delta] \cap \Gamma \lrcorner \hat{D}^\perp = [\hat{D}, \Delta] \cap \Gamma \lrcorner \hat{D}^\perp = \hat{D} \lrcorner \hat{D}^\perp \equiv 0.$$

From this result we gather that $A(\hat{D}^\perp) \in \mathcal{T}(\Delta; \Gamma)$ whenever Γ is involutive. The minimality of $\hat{D}$ implies $\hat{D} = A(\hat{D}^\perp)$, and from Theorem 8.3.1 we know that $A(\hat{D}^\perp)$ is involutive. □

Lemma 10.2.12 *If (Δ, Γ) forms an invariant pair, then $(\Delta, \hat{D})$ also forms an invariant pair.*

Proof: Since (Δ, Γ) form an invariant pair

$$[\Delta, \Gamma] \subset \Delta + \Gamma.$$

If we let $[\Delta, \hat{D}] = T$, then we have that $\hat{D} = T \cap \Gamma$ and

$$[\Delta, T \cap \Gamma] = T.$$

From the invariance condition, we get that

$$[\Delta, T \cap \Gamma] \subset [\Delta, \Gamma] \subset \Delta + \Gamma.$$

Forming the intersection, we get

$$[\Delta, T \cap \Gamma] \subset (\Delta + \Gamma) \cap T.$$

By definition, $\Delta \subset T$, so

$$[\Delta, T \cap \Gamma] \subset \Delta + T \cap \Gamma,$$

or

$$[\Delta, \hat{D}] \subset \Delta + \hat{D}.$$

□

We will define a controllability distribution to be any involutive distribution Γ which forms an invariant pair with Δ, and which is also the minimal element the collection $\mathcal{T}(\Delta;\Gamma)$ of distributions satisfying the conditions of Lemma 10.2.10.

If one tries to dualize the conditions in Lemma 10.2.10, the conditions obtained turn out to be difficult to work with. Instead, it is better to restate Lemma 10.2.10 in a slightly different form and then dualize the modified requirements.

Lemma 10.2.13 *Given any two distribution Δ and Γ, there exists a unique distribution $\hat{T}$ of minimal dimension which satisfies*

1. $\Delta \subset \hat{T}$.
2. $[\Delta, \hat{T} \cap \Gamma] = \hat{T}$.

Proof: Let $\hat{\mathcal{T}}(\Delta;\Gamma)$ denote the set of all distributions T satisfying

$$[\Delta, T \cap \Gamma] \subset T$$

This set is not empty since

$$[\Delta, V(M) \cap \Gamma] \subset V(M).$$

If $T_1, T_2 \in \hat{\mathcal{T}}(\Delta;\Gamma)$, then

$$\Delta \subset T_1, T_2 \Rightarrow \Delta \subset T_1 \cap T_2$$

and

$$\begin{aligned} [\Delta, T_1 \cap \Gamma] &\subset T_1 \\ [\Delta, T_2 \cap \Gamma] &\subset T_2 \\ \Rightarrow [\Delta, (T_1 \cap T_2) \cap \Gamma] &\subset T_1 \cap T_2. \end{aligned}$$

Since $\hat{\mathcal{T}}(\Delta;\Gamma)$ is nonempty and closed under smooth intersections, it must contain a unique minimal element.

To show equality, we begin by defining the distribution

$$K := [\Delta, \hat{T} \cap \Gamma].$$

This distribution is a subset of $\hat{T}$, so we must have that

$$[\Delta, K \cap \Gamma] \subset [\Delta, \hat{T} \cap \Gamma] = K.$$

This distribution also contains Δ, so it is a member of $K \in \hat{\mathcal{T}}(\Delta;\Gamma)$. From the minimality of $\hat{T}$, we conclude that $K = [\Delta, \hat{T} \cap \Gamma] = \hat{T}$. □

Note that the distributions $\hat{D}$ and $\hat{T}$ are related by $\hat{T} \cap \Gamma = \hat{D}$. In this modified form the conditions of this lemma can be dualized using Lemma 7.2.7 to require a unique maximal codistribution $\hat{T}^{\perp} \subset \Delta^{\perp}$ which satisfies

$$L_{\Delta} \hat{T}^{\perp} \subset \hat{T}^{\perp} + \Gamma^{\perp}. \tag{10.21}$$

10.3 Invariance and Dynamic State Feedback

In the previous section, we discussed an algorithm for finding the largest distribution which forms an invariant pair with a fixed distribution F and is contained in a fixed distribution Δ. In this section, we are going to examine how this distribution is affected if the control system corresponding to the distribution F is prolonged using the procedure described in Chapter 9.

To set this problem up, we will assume that we are given an s-dimensional distribution Δ, and an $(m+1)$-dimensional distribution F which are both defined over the $(n+1)$-dimensional state-time manifold $M \times \mathcal{R}$. Furthermore, we will assume that we are given a p-dimensional distribution $D \subset F \cap TM \times \{0\}$ which generates the prolongation of F. The state-time space for the prolonged system is the dense open subset of the bundle $G^{[D,F]}_{p+1}(M \times \mathcal{R})$ whose fibres over a point $(p,t) \in M \times \mathcal{R}$ consist of all the $(p+1)$-dimensional subspaces $S \in G^{[D,F]}_{k+1}(p,t)$ which satisfy $dt|_S \neq 0$. The prolonged distribution, which we will denote by F_D, is defined at each point $(p,\delta) \in G^{[D,F]}_{p+1}(M \times \mathcal{R})$ by the equation

$$F_D^{\perp}(p,\delta) = \pi^*(\delta^{\perp}).$$

Since the distribution Δ typically corresponds to the kernel of an output map

$$H : M \times \mathcal{R} \rightarrow N \times \mathcal{R}$$

and since the prolongation process induces a map

$$\hat{H} : G^{[D,F]}_{p+1}(M \times \mathcal{R}) \rightarrow N \times \mathcal{R}$$

defined by $\hat{H} := (H \circ \pi)$ from the prolonged space to the output space, a natural way to extend the distribution Δ to the prolonged space is to use the distribution $(\pi^*(\Delta^{\perp}))^{\perp}$ which, if Δ corresponds to the kernel of a map H, is equal to the kernel of the prolonged output map $\hat{H}$.

We want to try to find a relationship between the largest distribution contained in Δ and invariant with respect to F and the largest distribution contained in $(\pi^*(\Delta^{\perp}))^{\perp}$ and invariant with respect to F_D. As a first step toward this goal, we will compute one iteration of Algorithm 10.15 for each of the pairs (Δ, F) and $((\pi^*\Delta^{\perp})^{\perp}, F_d)$ and compare the resulting codistributions. We will work in a local coordinate system, and we will assume that Δ is involutive, that $\Delta \cap F$ has constant dimension, and that $\Delta \cap F \cap D$ has constant dimension. Under these assumptions, there exists a coordinate system

$$X : V \subset U \rightarrow \mathcal{R}^{n+1} \tag{10.22}$$

consisting of functions

$$h^1, \ldots, h^{r_h}, x^1, \ldots, x^{r_x}, \tilde{y}^1, \ldots, \tilde{y}^{r_{\tilde{y}}}, \hat{y}^1, \ldots, \hat{y}^{r_{\hat{y}}}, \tilde{z}^1, \ldots, \tilde{z}^{r_{\tilde{z}}}, \hat{z}^1, \ldots, \hat{z}^{r_{\hat{z}}}, t,$$

where $r_h + r_x + r_{\tilde{y}} + r_{\hat{y}} + r_{\tilde{z}} + r_{\hat{z}} = n$, which are selected selected such that

$$\Delta := \text{span}\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{r_x}}, \frac{\partial}{\partial \tilde{y}^1}, \ldots, \frac{\partial}{\partial \tilde{y}^{r_{\tilde{y}}}}, \frac{\partial}{\partial \hat{y}^1}, \ldots, \frac{\partial}{\partial \hat{y}^{r_{\hat{y}}}} \right\}$$

and such that, with respect to the ordered basis

$$\begin{array}{l} \frac{\partial}{\partial h^1}, \ldots, \frac{\partial}{\partial h^{r_h}}, \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{r_x}}, \frac{\partial}{\partial \tilde{y}^1}, \ldots, \frac{\partial}{\partial \tilde{y}^{r_{\tilde{y}}}}, \frac{\partial}{\partial \hat{y}^1}, \ldots, \frac{\partial}{\partial \hat{y}^{r_{\hat{y}}}}, \\ \frac{\partial}{\partial \tilde{z}^1}, \ldots, \frac{\partial}{\partial \tilde{z}^{r_{\tilde{z}}}}, \frac{\partial}{\partial \hat{z}^1}, \ldots, \frac{\partial}{\partial \hat{z}^{r_{\hat{z}}}}, \frac{\partial}{\partial t}, \end{array} \tag{10.23}$$

the following conditions are satisfied:

1. The distribution $\Delta \cap F \cap D$ is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 \\ G_2^2 \\ G_2^3 \\ I \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

2. The distribution $\Delta \cap F$ is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 & 0 \\ G_1^2 & G_2^2 \\ I & G_2^3 \\ 0 & I \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

3. The distribution D is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 & G_4^1 \\ G_2^2 & G_4^2 \\ G_2^3 & G_4^3 \\ I & 0 \\ 0 & G_4^5 \\ 0 & I \\ 0 & 0 \end{bmatrix}.$$

4. The distribution F is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 & 0 & G_3^1 & G_4^1 & F^1 \\ G_1^2 & G_2^2 & G_3^2 & G_4^2 & F^2 \\ I & G_2^3 & 0 & G_4^3 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & G_4^5 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{10.24}$$

Algorithm 10.15 is invariant under affine state feedback, so we can multiply the matrix 10.24 on the right by the feedback matrix

$$\begin{bmatrix} I & -G_2^3 & 0 & -G_4^3 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & -G_4^5 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

to get

$$\begin{bmatrix} 0 & 0 & G_3^1 & (G_4^1 - G_3^1 G_4^5) & F^1 \\ G_1^2 & (G_2^2 - G_1^2 G_2^3) & G_3^2 & (G_4^2 - G_1^2 G_4^3 - G_3^2 G_4^5) & F^2 \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The the vectors corresponding to the columns of this matrix can be supplemented with the vectors

$$\left\{ \frac{\partial}{\partial h^1}, \ldots, \frac{\partial}{\partial h^{r_h}}, \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{r_x}} \right\}$$

to form a basis of $T(M \times \mathcal{R})$ which can be represented by the matrix

$$\begin{bmatrix} I & 0 & 0 & 0 & G_3^1 & (G_4^1 - G_3^1 G_4^5) & F^1 \\ 0 & I & G_1^2 & (G_2^2 - G_1^2 G_2^3) & G_3^2 & (G_4^2 - G_1^2 G_4^3 - G_3^2 G_4^5) & F^2 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inverse of this matrix is

$$\begin{bmatrix} I & 0 & 0 & 0 & -G_3^1 & (G_3^1 G_4^5 - G_4^1) & -F^1 \\ 0 & I & -G_1^2 & (G_1^2 G_2^3 - G_2^2) & -G_3^2 & (G_1^2 G_4^3 + G_3^2 G_4^5 - G_4^2) & -F^2 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and its rows can be viewed as one-forms expressed with respect to the standard dual basis of 10.23. The first two blocks of rows in this matrix span $F^\perp$, the first, fourth, fifth, and sixth blocks of rows span $\Delta^\perp$, and the first block of rows spans $F^\perp \cap \Delta^\perp$.

If we set $J_0 = \Delta^\perp$, then the first iteration of Algorithm 10.15 is described by

$$J_1 = J_0 + L_F(J_0 \cap F^\perp).$$

In local coordinates, we have that

$$J_0 = \text{span}\{dh^1, \ldots, dh^{r_h}, d\tilde{z}^1, \ldots, d\tilde{z}^{r_{\tilde{z}}}, d\hat{z}^1, \ldots, d\hat{z}^{r_{\hat{z}}}, dt\},$$

and using the procedure presented in the proof of Theorem 10.2.2 we can compute that

$$J_1 = \text{span}\{dh^1, \ldots, dh^{r_h}, d\tilde{z}^1, \ldots, d\tilde{z}^{r_{\tilde{z}}}, d\hat{z}^1, \ldots, d\hat{z}^{r_{\hat{z}}}, dt, dG_3^1, d(G_3^1 G_4^5 - G_4^1), dF^1\}$$

where we use terms like dG_3^1 to denote the differentials of each of the elements in the matrix G_3^1.

Having obtained J_1 for the unprolonged system, we turn our attention to the prolonged system. The prolongation process is equivalent to appending integrators onto the inputs associated with the first and third blocks of columns in the matrix 10.24. We can form a new local coordinate system on the prolonged state-time space by adding the coordinates $v^1, \ldots, v^{r_{\tilde{y}}}, u^1, \ldots, u^{r_{\hat{z}}}$ to the coordinate set 10.22. With respect to these coordinates, the distribution F_D of the prolonged system is spanned by the columns of the matrix

$$\begin{bmatrix} 0 & G_4^1 & 0 & 0 & G_3^1 U + F^1 \\ G_2^2 & G_4^2 & 0 & 0 & G_1^2 V + G_3^2 U + F^2 \\ G_2^3 & G_4^3 & 0 & 0 & V \\ 0 & G_4^5 & 0 & 0 & U \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (10.25)$$

which is expressed relative to the ordered basis

$$\text{span}\left\{\frac{\partial}{\partial h}, \frac{\partial}{\partial x}, \frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \tilde{z}}, \frac{\partial}{\partial \hat{y}}, \frac{\partial}{\partial \hat{z}}, \frac{\partial}{\partial v}, \frac{\partial}{\partial u}, \frac{\partial}{\partial t}\right\}$$

where we use terms like $\frac{\partial}{\partial h}$ to mean

$$\frac{\partial}{\partial h} = \left\{\frac{\partial}{\partial h^1}, \ldots, \frac{\partial}{\partial h^{r_h}}\right\}.$$

As before, we can append the vectors

$$\left\{\frac{\partial}{\partial h}, \frac{\partial}{\partial x}, \frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \tilde{z}}\right\}$$

to the columns of the matrix 10.25 to form a basis of the tangent space of the prolonged state-time manifold. This basis can be represented by the matrix

$$\begin{bmatrix} I & 0 & 0 & 0 & 0 & G_4^1 & 0 & 0 & G_3^1 U + F^1 \\ 0 & I & 0 & 0 & G_2^2 & G_4^2 & 0 & 0 & G_1^2 V + G_3^2 U + F^2 \\ 0 & 0 & I & 0 & G_2^3 & G_4^3 & 0 & 0 & V \\ 0 & 0 & 0 & I & 0 & G_4^5 & 0 & 0 & U \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inverse of this matrix is

$$\begin{bmatrix} I & 0 & 0 & 0 & 0 & -G_4^1 & 0 & 0 & -G_3^1 U - F^1 \\ 0 & I & 0 & 0 & -G_2^2 & -G_4^2 & 0 & 0 & -G_1^2 V - G_3^2 U - F^2 \\ 0 & 0 & I & 0 & -G_2^3 & -G_4^3 & 0 & 0 & -V \\ 0 & 0 & 0 & I & 0 & -G_4^5 & 0 & 0 & -U \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Identifying the rows of this matrix with differential one-forms expressed relative to the standard cobasis for our local coordinates, we find that the first four blocks of rows span $F_D^\perp$, the first, fourth, sixth, and ninth blocks of rows span the codistribution $K_0 = \pi^*(J_0)$, and the first and fourth blocks of rows span $K_0 \cap F_D^\perp$. computing $K_1 = K_0 + L_{F_D}(K_0 \cap F_D^\perp)$ we find that

$$K_1 = \text{span}\{dh, d\tilde{z}, d\hat{z}, dt, dU, dG_4^5, dG_4^1, d(G_3^1 U + F^1)\}.$$

If we carry out the indicated differentiations, we can rewrite J_1 as

$$J_1 = \text{span}\{dh, d\tilde{z}, d\hat{z}, dt, dG_3^1, (G_3^1 dG_4^5 - dG_4^1), dF^1\}$$

and K_1 as

$$K_1 = \text{span}\{dh, d\tilde{z}, d\hat{z}, dt, dU, dG_4^5, dG_4^1, (dG_3^1 U + dF^1)\}.$$

From these expressions, it is clear that the relationship between K_1 and J_1 is dependent on the functions which determine the G matrices. For example, if G_3^1 is constant, then $\pi^* J_1 \subset K_1$. If, on the other hand, G_4^5 is constant, then $K_1 \cap (\ker \pi)^\perp \subset J_1$. In particular, note that if we set $D = \Delta \cap F$, then this latter case will always occur. In fact we can show that this relationship holds for each of the corresponding pairs J_i and K_i in the filtrations generated by Algorithm 10.15.

Lemma 10.3.1 *Given an s-dimensional distribution Δ, and an $(m+1)$-dimensional distribution F which are both defined over the $(n+1)$-dimensional state-time manifold $M \times \mathcal{R}$, assume that Δ is involutive, and that $\Delta \cap F$ has constant dimension p Let $D = \Delta \cap F$, and let F_D denote the prolonged distribution generated by D. Let $J_0 = \Delta^\perp$ and $K_0 = \pi^* J_0$. If the filtrations generated by applying Algorithm 10.15 respectively to the pairs (J_0, F) and (K_0, F_D) are both constant dimensional on a neighborhood of the prolonged state-time space, and the codistributions $F^\perp + J_i$ all have constant dimension on this neighborhood, then the filtrations satisfy*

$$\begin{array}{ccccccccc} \pi^* J_0 & \subset & \pi^* J_1 & \subset & \cdots & \subset & \pi^* J_i & \subset & \cdots \\ \cup & & \cup & & & & \cup & & \\ K_0 \cap (\ker \pi)^\perp & \subset & K_1 \cap (\ker \pi)^\perp & \subset & \cdots & \subset & K_i \cap (\ker \pi)^\perp & \subset & \cdots \end{array}$$

In particular, $\hat{K} \cap (\ker \pi)^\perp \subset \pi^* \hat{J}$.

Proof: Corresponding to the filtration of codistributions

$$J_0 \subset J_1 \subset \cdots \subset J_i \subset \cdots,$$

we have the filtration of distributions

$$\hat{J}^\perp \subset \cdots \subset J_i^\perp \subset \cdots \subset J_0^\perp = \Delta.$$

Let I be the lowest integer such that $J_I = \hat{J}$. For simplicity, we will assume $I = 2$, but the proof of the general case will proceed in exactly the same way. Because of the constant dimension assumptions, there must exist a set of coordinates

$$\tilde{h}^1, .., \tilde{h}^{r_{\tilde{h}}}, \hat{h}^1, .., \hat{h}^{r_{\hat{h}}}, \tilde{x}^1, .., \tilde{x}^{r_{\tilde{x}}}, \hat{x}^1, .., \hat{x}^{r_{\hat{x}}}, \tilde{y}^1, .., \tilde{y}^{r_{\tilde{y}}}, \hat{y}^1, .., \hat{y}^{r_{\hat{y}}}, \tilde{z}^1, .., \tilde{z}^{r_{\tilde{z}}}, \hat{z}^1, .., \hat{z}^{r_{\hat{z}}}, t,$$

where $r_{\tilde{h}} + r_{\hat{h}} + r_{\tilde{x}} + r_{\hat{x}} + r_{\tilde{y}} + r_{\hat{y}} + r_{\tilde{z}} + r_{\hat{z}} = n$, which are selected selected such that

$$J_2^\perp = \operatorname{span}\left\{\frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \hat{y}}\right\},$$

$$J_1^\perp = \operatorname{span}\left\{\frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \hat{y}}, \frac{\partial}{\partial \tilde{z}}, \frac{\partial}{\partial \hat{z}}\right\},$$

$$J_0^\perp = \operatorname{span}\left\{\frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \hat{y}}, \frac{\partial}{\partial \tilde{z}}, \frac{\partial}{\partial \hat{z}}, \frac{\partial}{\partial \tilde{x}}, \frac{\partial}{\partial \hat{x}}\right\},$$

and such that F is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 & 0 & 0 & G_4^1 & F^1 \\ 0 & 0 & G_3^2 & G_4^2 & F^2 \\ 0 & G_2^3 & G_3^3 & G_4^3 & F^3 \\ G_1^4 & G_2^4 & G_3^4 & G_4^4 & F^4 \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

which is written relative to the ordered basis

$$\frac{\partial}{\partial \tilde{h}}, \frac{\partial}{\partial \tilde{x}}, \frac{\partial}{\partial \tilde{z}}, \frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \hat{y}}, \frac{\partial}{\partial \hat{z}}, \frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial \hat{h}}, \frac{\partial}{\partial t}.$$

Applying Algorithm 10.15 we find that

$$\begin{aligned}
J_0 &= \operatorname{span}\{d\tilde{h}, d\hat{h}, dt\} \\
J_1 &= \operatorname{span}\{d\tilde{h}, d\hat{h}, dt, dG_4^1, dF^1\} \\
&= \operatorname{span}\{d\tilde{h}, d\hat{h}, d\tilde{x}, d\hat{x}, dt\} \\
J_2 &= \operatorname{span}\{d\tilde{h}, d\hat{h}, dt, dG_4^1, dF^1, dG_3^2, dG_4^2, dF^2\} \\
&= \operatorname{span}\{d\tilde{h}, d\hat{h}, d\tilde{x}, d\hat{x}, d\tilde{z}, d\hat{z}, dt\}. \\
J_3 &= \operatorname{span}\{d\tilde{h}, d\hat{h}, dt, dG_4^1, dF^1, dG_3^2, dG_4^2, dF^2, dG_2^3, dG_3^3, dG_4^3, dF^3\} \\
&= \operatorname{span}\{d\tilde{h}, d\hat{h}, d\tilde{x}, d\hat{x}, d\tilde{z}, d\hat{z}, dt\}.
\end{aligned}$$

Prolonging F about $D = \Delta \cap F$ produces the distribution F_D which can be represented by the matrix of functions

$$\begin{bmatrix}
0 & 0 & 0 & 0 & U \\
0 & 0 & 0 & 0 & G_4^1 U + F^1 \\
0 & 0 & G_3^2 & 0 & G_4^2 U + F^2 \\
0 & G_2^3 & G_3^3 & 0 & G_4^3 U + F^3 \\
G_1^4 & G_2^4 & G_3^4 & 0 & G_4^4 U + F^4 \\
I & 0 & 0 & 0 & 0 \\
0 & I & 0 & 0 & 0 \\
0 & 0 & I & 0 & 0 \\
0 & 0 & 0 & I & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

which is written relative to the ordered basis

$$\frac{\partial}{\partial \hat{h}} \frac{\partial}{\partial \tilde{h}}, \frac{\partial}{\partial \tilde{x}}, \frac{\partial}{\partial \tilde{z}}, \frac{\partial}{\partial \tilde{y}}, \frac{\partial}{\partial \hat{y}}, \frac{\partial}{\partial \hat{z}}, \frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial u}, \frac{\partial}{\partial t}.$$

For each J_i, we will now compute the codistribution $\pi^* J_i + L_{F_D}(\pi^* J_i \cap F_D^\perp)$.

$$\begin{aligned}
\Omega_1 &= \pi^* J_0 + L_{F_D}(\pi^* J_0 \cap F_D^\perp) \\
&= \{d\hat{h}, d\tilde{h}, dt, d(G_4^1 U + F^1), dU\} \\
\Omega_2 &= \pi^* J_1 + L_{F_D}(\pi^* J_1 \cap F_D^\perp) \\
&= \{d\hat{h}, d\tilde{h}, d\hat{x}, d\tilde{x}, dt, dG_3^2, d(G_4^1 U + F^1), d(G_4^2 U + F^2), dU\} \\
&= \{d\hat{h}, d\tilde{h}, dt, dG_4^1, dF^1, dG_3^2, d(G_4^2 U + F^2), dU\} \\
\Omega_3 &= \pi^* J_2 + L_{F_D}(\pi^* J_2 \cap F_D^\perp) \\
&= \{d\hat{h}, d\tilde{h}, d\hat{x}, d\tilde{x}, d\tilde{z}, d\hat{z}, dt, dG_2^3, dG_3^3, dG_3^2, d(G_4^1 U + F^1), \\
&\quad d(G_4^2 U + F^2), d(G_4^3 U + F^3), dU\} \\
&= \{d\hat{h}, d\tilde{h}, dt, dG_4^1, dF^1, dG_3^2, dG_4^2, dF^2, dG_2^3, dG_3^3, dG_3^2, d(G_4^3 U + F^3), dU\}
\end{aligned}$$

From these computations, it is clear that

$$\begin{aligned}
\Omega_1 \cap (\ker \pi)^\perp &\subset \pi^* J_1 \\
\Omega_2 \cap (\ker \pi)^\perp &\subset \pi^* J_2 \\
\Omega_3 \cap (\ker \pi)^\perp &\subset \pi^* J_3
\end{aligned}$$

Furthermore, since $F_D^\perp \subset (\ker \pi)^\perp$, we must also have that

$$\begin{aligned}
\Omega_1 \cap F_D^\perp &\subset \pi^* J_1 \cap F_D^\perp \\
\Omega_2 \cap F_D^\perp &\subset \pi^* J_2 \cap F_D^\perp \\
\Omega_3 \cap F_D^\perp &\subset \pi^* J_3 \cap F_D^\perp.
\end{aligned}$$

Since $\Omega_1 = K_1$, we obviously have that $K_1 \cap (\ker \pi)^\perp \subset \pi^* J_1$ and that $K_1 \cap F_D^\perp \subset \pi^* J_1 \cap F_D^\perp$. Assume that for some $i \geq 1$, $K_i \cap F_D^\perp \subset \pi^* J_i \cap F_D^\perp$. Expanding out the expression for K_{i+1}, we find that

$$\begin{aligned} K_{i+1} &= K_i + L_{F_D}(K_i \cap F_D^\perp) \\ &= K_0 + L_{F_D}(K_0 \cap F_D^\perp) + L_{F_D}(K_1 \cap F_D^\perp) + \ldots + L_{F_D}(K_i \cap F_D^\perp) \\ &= J_0 + L_{F_D}(K_i \cap F_D^\perp) \end{aligned}$$

Using the induction hypothesis, we find that

$$K_{i+1} \subset J_0 + L_{F_D}(J_i \cap F_D^\perp) \subset J_i + L_{F_D}(J_i \cap F_D^\perp) = \Omega_{i+1}$$

which implies that $K_{i+1} \cap (\ker \pi)^\perp \subset J_{i+1}$. □

Based on Lemma 10.3.1 it appears that a second prolongation might result in another layer of nested codistributions. However, if we were to compute the codistribution $K_0^\perp \cap F_D$ on the prolonged space, we would find that is equal to $F_D \cap TM \times \{0\}$. This distribution contains all the controls, so the prolongation which it generates is equivalent to F_D. Consequently, nothing is gained. However, if we begin with the codistribution K_1, then $K_1 \cap F_D$ will generally be a proper subset of the controls and there may be something to be gained from the prolongation. To formalize this procedure we can define a prolongation algorithm.

Algorithm 10.3.1 *Given a distribution Δ and a distribution F which are both defined over a state-time space $M \times \mathcal{R}$, define $K_0^0 = \Delta^\perp$, $F_0 = F$, and $M_0 \times \mathcal{R} = M \times \mathcal{R}$. At the ith step of the algorithm:*

1. *Compute $D_i = (K_{i-1}^{i-1})^\perp \cap F_{i-1}$. Let $p_i = dim(D_i)$.*
2. *Form the prolongation of F_{i-1} generated by D_i. Let F_i denote the distribution associated with this prolongation.*
3. *Set $M_i \times \mathcal{R} = G_{p_i+1}^{[D_i, F_{i-1}]}(M_{i-1} \times \mathcal{R})$, and and let π_i denote the projection $\pi_i : M_i \times \mathcal{R} \to M_{i-1} \times \mathcal{R}$.*
4. *Define $K_{i-1}^i = \pi_i^* K_{i-1}^{i-1}$.*
5. *Compute $K_i^i = K_{i-1}^i + L_{F_i}\left(K_{i-1}^i \cap F_i^\perp\right)$.*

At each step of this algorithm, Lemma 10.3.1 ensures that

$$\begin{array}{lllllllll} \pi_i^* K_{i-1}^{i-1} & \subset & \pi_i^* K_i^{i-1} & \subset & \cdots & \subset & \pi_i^* K_k^{i-1} & \subset & \cdots \\ \cup & & \cup & & & & \cup & & \\ K_{i-1}^i \cap (\ker \pi_i)^\perp & \subset & K_i^i \cap (\ker \pi_i)^\perp & \subset & \cdots & \subset & K_k^i \cap (\ker \pi_i)^\perp & \subset & \cdots . \end{array}$$

Consequently, if each of the filtrations

$$K_i^i \subset K_{i+1}^i \subset K_{i+2}^i \subset \cdots$$

stabilizes after a finite number of iterations, then Algorithm 10.3.1 will also stabilize after a finite number of iterations. This algorithm will prove useful when we look at the disturbance decoupling and noninteracting controls problems. It is closely related to a procedure called Singh's algorithm, and the interested reader may want to consult [11] for a good treatment of this topic.

Chapter 11

The Disturbance Decoupling and Noninteracting Control Problems

In this chapter, we will apply the material from Chapter 10 to the disturbance decoupling and noninteracting control problems.

The first section of this chapter discusses the disturbance decoupling problem. This problem arises when a control system is subjected to external disturbance inputs. In such cases, it is desirable to minimize the influence of these disturbances on the outputs of the system. Therefore, the disturbance decoupling problem is posed by asking when a feedback transformation exists such that, in the transformed system, the disturbance inputs do not influence the system outputs.

The second section of this chapter discusses the noninteracting control problem. This problem arises in situations where it is conceptually or practically advantageous to treat a multi-input multi-output system as if it consisted of a collection of disconnected subsystems. Since this is not always possible, it is useful to have a test which allow the controls engineer to determine if a feedback transformation exists which will will decouple the input-output behavior of the system.

11.1 Disturbance Decoupling

This section discusses the disturbance decoupling problem. To set this problem up, we will consider an affine control system which is defined by a distribution F over a smooth state-time manifold $M \times \mathcal{R}$. We will assume that F is defined as the sum of two distributions $F = F_u + F_d$. The distribution F_u is spanned by vectors of the form

$$f_e = \begin{bmatrix} f \\ 1 \end{bmatrix} \quad g_{ei} = \begin{bmatrix} g_i \\ 0 \end{bmatrix}$$

which represent the drift vector field and the m inputs which can be used to control the system. The distribution F_d is spanned by s vectors of the form

$$r_{ei} = \begin{bmatrix} r_i \\ 0 \end{bmatrix}$$

which represent disturbance inputs over which we have no control. With respect to a local coordinate chart (x, U), this control system can then be represented by the equation

$$\dot{x}^i = f_e^i + \sum_{j=1}^{m} g_{ej}^i u^j + \sum_{k=1}^{s} r_{ek} v^k.$$

We will call the inputs $v^1, \ldots, v^s$ the disturbance inputs.

Given a set of outputs functions $y^1 = h^1(q), \ldots, y^p = h^p(q)$, we will use $y^i(q_0, t_0, u, v)$ to denote the time function which results when the system is started with initial condition $(q_0, t_0) \in M \times \mathcal{R}$ and driven by the input function $u(t)$ and the disturbance function $v(t)$. A disturbance input $v(t)$ is said to be decoupled from the output y^i if and only if for every $(q_0, t_0) \in M \times \mathcal{R}$, every $u(t) \in C^\infty(\mathcal{R}, \mathcal{R}^m)$, and every $v(t) \in C^\infty(\mathcal{R}, \mathcal{R}^m)$, $y^i(q_0, t_0, u, 0) = y^i(q_0, t_0, u, v)$. In order to study this problem, we need to define the codistribution

$$\begin{aligned} \mathcal{O}_{fg} &= \operatorname{span}\left\{ d\lambda \in T^*M \mid \lambda = L_{X_1} L_{X_2} \cdots L_{X_k} h^j \;\; \forall k \geq 0,\; X_k \in \{f_e, g_{e1}, \ldots, g_{em}\} \right\} \\ &+ \operatorname{span}\{dt\}. \end{aligned}$$

which is called the observability distribution associated with the control system defined by f_e, g_{e1}, $\ldots$, g_{em}, and $h^1, \ldots, h^p$. The fg subscript is used to emphasize that this codistribution is dependent on the choice of these vector fields and that this codistribution is invariant with respect to these vector fields. With this setup, we now state a lemma which gives a necessary condition for an input to be decoupled from an output.

Lemma 11.1.1 *If a disturbance input $v^i(t)$ is decoupled from an output $y^j = h^j(p, t)$, then the corresponding vector field r_{ei} lies in the perp of the codistribution $\mathcal{O}_{fg}$*

Proof: Let v and $\tilde{v}$ be two arbitrary disturbance inputs and let $y^j(q_0, t_0, u, v)$ and $y^j(q_0, t_0, u, \tilde{v})$ denote the corresponding outputs. Since the disturbances are assumed to be decoupled from the outputs, these functions and all their time derivatives must be identical. The time derivate of $y^j(q_0, t_0, u, v)$ is equal to the Lie derivate of h^j with respect to the vector field $f_e + g_{ei} u^i(t) + r_{ek} v^k(t)$ which is defined over $M \times \mathcal{R}$. Similarly, the time derivative of $y^j(q_0, t_0, u, \tilde{v})$ is equal to the Lie derivative of h^j with respect to the vector field $f_e + g_{ei} u^i(t) + r_{ek} \tilde{v}^k(t)$. This implies that

$$\frac{dy^j}{dt}(q_0, t_0, u, v) = L_{f_e} h^j + L_{g_{ei}} h^j u^i(t) + L_{r_{ek}} h^j v^k(t).$$

Since the functions $v^k(t)$ are arbitrary, and this function is assumed to be independent of the disturbance inputs, we must have $L_{r_{ek}} h^j = r_{ek} \rfloor dh^j \equiv 0$. Taking the second derivative of $y^j(q_0, t_0, u, v)$

$$\begin{aligned}\frac{d^2 y^j}{dt^2}(q_0, t_0, u, v) &= L_{f_e}^2 h^j + L_{f_e} L_{g_{ei}} h^j u^i(t) + L_{g_{ej}} L_{f_e} h^j u^j(t) + L_{g_{ej}} L_{g_{ei}} h^j u^i(t) u^j(t) \\ &+ L_{g_{ei}} h^j \dot{u}^i(t) + L_{r_{ek}} L_{f_e} h^j v^k(t) + L_{r_{ek}} L_{g_{ei}} h^j u^i(t) v^k(t).\end{aligned}$$

Since the value of the second derivative is also assumed to be independent of the disturbance inputs, and this condition is true for arbitrary $u^i(t)$, we must also have $L_{r_{ek}} L_{f_e} h^j = r_{ek} \rfloor dL_{f_e} h^j \equiv 0$ and $L_{r_{ek}} L_{g_{ei}} h^j = r_{ek} \rfloor dL_{g_{ei}} h^j \equiv 0$.

By continuing to take higher order derivatives, we obtain the condition that $r_{ek} \rfloor d\lambda \equiv 0$ for any $\lambda = L_{X_1} L_{X_2} \cdots L_{X_k} h^j$ where $k \geq 0$ and $X_i \in \{f_e, g_{e1}, \ldots, g_{em}\}$. □

11.1.1 Disturbance Decoupling Using Static State Feedback

The problem which we will consider in this section is to determine whether or not there exists an affine feedback transformation such that, in the transformed system, the disturbance inputs are decoupled from the outputs. Define

$$J = \text{span}\{dh^1, \ldots, dh^p, dt\},$$

and let $\hat{J}$ denote the smallest codistribution which contains J and forms an invariant pair with F. The following theorem provides necessary and sufficient conditions for the solvability of this problem when $\hat{J}$ has constant dimension.

Theorem 11.1.1 *If the codistribution $\hat{J}$ has constant dimension, then the disturbance decoupling problem is solvable if and only if $F_d \subset \hat{J}^\perp$.*

Proof: If the disturbance inputs are decoupled from the outputs, then from Lemma 11.1.1, $r_{ei} \in \mathcal{O}_{fg}^\perp$. Furthermore,

$$\begin{aligned}L_{f_e} \mathcal{O}_{fg} &\subset \mathcal{O} \\ L_{g_{ei}} \mathcal{O}_{fg} &\subset \mathcal{O},\end{aligned}$$

so $L_F(F^\perp \cap \mathcal{O}_{fg}) \subset \mathcal{O}_{fg}$. Therefore, $(\mathcal{O}_{fg}^\perp, F)$ is an invariant pair. The codistribution $\mathcal{O}_{fg}$ also contains J. Consequently, $\hat{J} \subset \mathcal{O}_{fg}$. These facts imply that $r_{ei} \in \mathcal{O}_{fg}^\perp \subset \hat{J}^\perp$.

Conversely, if each $r_{ei} \in \hat{J}^\perp$, then Lemma 10.1.5 says that there exists a coordinate system and affine state feedback such that the control system can be represented in the normal form

$$\begin{aligned}\dot{z}^1 &= f^1(z^1, z^2) + \sum_{i=1}^m g_i^1(z^1, z^2) u^i + \sum_{k=1}^s r_k^1(z^1, z^2) v^k \\ \dot{z}^2 &= f^2(z^2) + \sum_{i=1}^m g_i^2(z^2) u^i \\ y^j &= h^j(z^2).\end{aligned}$$

Written in this form, it is clear that the outputs are not influenced by the disturbance inputs, so this choice of feedback solves the disturbance decoupling problem. □

11.1.2 Disturbance Decoupling Using Dynamic State Feedback

Even if the disturbance decoupling problem is not solvable using static state feedback, it may still be possible to decouple the disturbance inputs using dynamic state feedback. This possibility is based on the discussion in Section 10.3 which showed that by applying Algorithm 10.3.1 the size of the largest invariant codistribution which contains J may be decreased on the prolonged space. To test whether F_d can be decoupled using dynamic state feedback, the following procedure can be used.

1. Set $K_0^0 = J$, $F_0 = F$, and $i = 0$.
2. If $F_d \not\subset (K_i^i)^\perp$ then the disturbance cannot be decoupled using further bundle prolongations.
3. If $F_d \subset (K_i^i)^\perp$, construct the filtration

$$K_i^i \subset K_{i+1}^i \subset \cdots \subset \hat{K}^i$$

   using Algorithm 10.15.

4. If $F_d \subset (\hat{K}^i)^\perp$, then F_d can be decoupled using static state feedback on the system F_i defined over $M_i \times \mathcal{R}$.
5. If $F_d \not\subset (\hat{K}^i)^\perp$, then set $i = i + 1$, perform one iteration of Algorithm 10.3.1, and return to step 2.

11.2 Noninteracting Controls

In this section, we will consider the block input-output decoupling problem. This problem can be stated as follows. Given an affine system which is described by a distribution F defined over the state-time manifold $M \times \mathcal{R}$ and a collection of output blocks H_1 to H_k of the form $H_i = (h_i^1, \ldots, h_i^{p_i}, t)$, determine if there exists a decomposition of F of the form

$$\begin{bmatrix} G^1 & \cdots & G^k & F \\ 0 & & 0 & 1 \end{bmatrix}$$

and a state feedback transformation such that, in the transformed system, the ith output block is influenced by the ith input block and is decoupled from the jth input block for $j \neq i$. If the control system is strongly accessible, then the following theorem provides a necessary and sufficient condition for the existence of such a decomposition.

Theorem 11.2.1 *Let J_i be the smallest codistribution which contains dH_i and forms an invariant pair with F. If the control system corresponding to F is strongly accessible, then the block input-output decoupling problem is solvable if and only if each J_i satisfies*

$$J_i \cap (\sum_{s \neq i} J_s) = span\{dt\}. \tag{11.1}$$

Proof: To prove the sufficiency of this theorem, we begin by noting that each of the codistributions J_i is involutive, so there exist smooth functions $\lambda_i^1, \ldots, \lambda_i^{r_i}$ such that

$$J_i = \text{span}\{d\lambda_i^1, \ldots, d\lambda_i^{r_i}, dt\}.$$

If we define the codistributions $K_i := \sum_{s \neq i} J_s$, then condition 11.1 can be rewritten as

$$J_i \cap K_i = \text{span}\{dt\}, \;\; 1 \leq i \leq k.$$

Using this fact, it can be shown that for each $i \in \{1, \ldots, k\}$,

$$K_1 \cap K_2 \cap \cdots \cap K_{i-1} = J_{i+1} + \cdots + J_k.$$

From this, it is easy to show that

$$K_1 \cap K_2 \cap \cdots \cap K_{i-1} = K_1 \cap K_2 \cap \cdots \cap K_i + J_i. \tag{11.2}$$

Suppose the codistribution $K_1 \cap K_2 \cap \cdots \cap K_i$ is spanned by a set of differentials

$$K_1 \cap K_2 \cap \cdots \cap K_i = \text{span}\{d\beta^1, \ldots, d\beta^s\}.$$

Since $K_i \cap J_i = \text{span}\{dt\}$, the equation $K_1 \cap K_2 \cap \cdots \cap K_i \cap J_i = \text{span}\{dt\}$ also holds, and this fact together with equation 11.2 implies that the codistribution $K_1 \cap K_2 \cap \cdots \cap K_{i-1}$ will be spanned by the differentials

$$K_1 \cap K_2 \cdots \cap K_{i-1} = \text{span}\{d\beta^1, \ldots, d\beta^s, d\lambda_i^1, \ldots, d\lambda_i^{r_i}, dt\}.$$

Therefore, starting with $i = k$ and working backwards recursively, we find that K_1 is spanned by the differentials

$$K_1 = \text{span}\{d\lambda_2^1, \ldots, d\lambda_2^{r_2}, d\lambda_3^1, \ldots, d\lambda_3^{r_3}, \ldots, d\lambda_k^1, \ldots, d\lambda_k^{r_k}, dt\}$$

and that

$$J = K_1 + J_1 = \text{span}\{d\lambda_1^1, \ldots, d\lambda_1^{r_1}, d\lambda_2^1, \ldots, d\lambda_2^{r_2}, \ldots, d\lambda_k^1, \ldots, d\lambda_k^{r_k}, dt\}.$$

Consequently, all of the functions λ_i^j can be taken together with t and an additional set of linearly independent function $x^1, \ldots, x^q$ to form a local coordinate chart on $T^*(M \times \mathcal{R})$.

Since each of the distributions $J_i^\perp$ forms an invariant pair with F, we have that

$$C(J_i^\perp \cap F^\perp) \subset J_i.$$

Therefore, Theorem 8.3.2 ensures that locally there exists a basis $\omega^1, \ldots, \omega^{p_i}$ for $J_i \cap F^\perp$ such that each ω^k can be expressed as

$$\omega^k = \sum_{j=1}^{r_i} a_j(\lambda_i) d\lambda_i^j.$$

Since this is true for each J_i, there exists a basis for $F^\perp$ of the form

$$\begin{bmatrix} I & 0 & 0 & 0 & 0 & G_2^1(\lambda_1, t) & 0 & 0 & F^1(\lambda_1, t) \\ 0 & \ddots & 0 & 0 & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & I & 0 & 0 & 0 & 0 & G_{k+1}^k(\lambda_k, t) & F^k(\lambda_k, t) \\ 0 & 0 & 0 & I & G_1^{k+1}(\lambda, x, t) & G_2^{k+1}(\lambda, x, t) & \cdots & G_{k+1}^{k+1}(\lambda, x, t) & F^{k+1}(\lambda, x, t) \end{bmatrix}$$

expressed relative to the ordered cobasis

$$d\hat{\lambda}_1, d\hat{\lambda}_2, \ldots, d\hat{\lambda}_k, d\hat{x}, d\tilde{\lambda}_1, d\tilde{\lambda}_2, \ldots, d\tilde{\lambda}_k, d\tilde{x}, dt$$

where we have divided each set of coordinate functions into two blocks, $d\tilde{\lambda}_i$ and $d\hat{\lambda}_i$ and where terms like dx denote the set $\{dx^1, \ldots, dx^q\}$.

Furthermore this can be extended to a cobasis of the form

$$\begin{bmatrix} I & \cdots & 0 & 0 & 0 & G_2^1(\lambda_1, t) & \cdots & 0 & F^1(\lambda_1, t) \\ 0 & \ddots & 0 & 0 & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & I & 0 & 0 & 0 & 0 & G_{k+1}^k(\lambda_k, t) & F^k(\lambda_k, t) \\ 0 & 0 & 0 & I & G_1^{k+1}(\lambda, x, t) & G_2^{k+1}(\lambda, x, t) & \cdots & G_{k+1}^{k+1}(\lambda, x, t) & F^{k+1}(\lambda, x, t) \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Inverting this matrix, we set that F is spanned by the columns of the matrix of functions

$$\begin{bmatrix} 0 & -G_2^1(\lambda_1, t) & 0 & 0 & -F^1(\lambda_1, t) \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & -G_{k+1}^k(\lambda_k, t) & -F^k(\lambda_k, t) \\ -G_1^{k+1}(\lambda, x, t) & -G_2^{k+1}(\lambda, x, t) & \cdots & -G_{k+1}^{k+1}(\lambda, x, t) & -F^{k+1}(\lambda, x, t) \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{11.3}$$

Written in this form, it is clear that the inputs can be grouped into blocks such that the noninteraction conditions are satisfied. Assuming the input and drift vector fields for the uncompensated system are known, a feedback which renders the system noninteractive will be given by the feedback transformation which maps the matrix corresponding to the uncompensated system onto the matrix 11.3. Therefore, the problem is solvable.

In order to prove necessity, we will assume that the problem is solvable and will let D_i denote the distribution spanned by the ith control block. The noninteracting conditions require that D_i be decoupled from each block of outputs H_j for $j \neq i$. From Theorem 11.1.1, we know that a necessary condition for this to occur is that D_i must be a subset of the largest controlled invariant distribution contained in $\ker(\sum_{j \neq i} dH_j)$, which we will denote by I_i. Since this codistribution is involutive, there exists a set of differentials defined such that

$$I_i = \{d\tilde{\lambda}, d\hat{\lambda}, dt\}.$$

Since $I_i^\perp$ forms an invariant pair with F, $F^\perp$ can be represented by the first two blocks of rows in a matrix of the form

$$\begin{bmatrix} I & 0 & 0 & G_2^1(\lambda,t) & F^1(\lambda,t) \\ 0 & I & G_1^2(\lambda,x,t) & G_2^2(\lambda,x,t) & F^2(\lambda,x,t) \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

which is expressed relative to the ordered cobasis

$$d\hat{\lambda}, d\hat{x}, d\tilde{x}, d\tilde{\lambda}, dt.$$

Furthermore, if we compute the largest controllability distribution contained in $I_i^\perp$, which we will denote by M_i, then there exists a set of coordinates λ, x, z, t such that

$$M_i^\perp = \text{span}\{d\lambda, dx, dt\},$$

and we can further refine this description to

$$\begin{bmatrix} I & 0 & 0 & 0 & G_2^1(\lambda,t) & F^1(\lambda,t) \\ 0 & I & 0 & 0 & G_2^2(\lambda,x,t) & F^2(\lambda,x,t) \\ 0 & 0 & I & G_1^2(\lambda,x,z,t) & G_2^3(\lambda,x,z,t) & F^3(\lambda,x,z,t) \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

expressed relative to the ordered cobasis

$$d\hat{\lambda}, d\hat{x}, d\hat{z}, d\tilde{z}, d\tilde{x}, d\tilde{\lambda}, dt.$$

Since the system is assumed to be decoupled, we must have that $J_i \cap M_i^\perp = \{dt\}$. If this were not the case, then there would exist an $\omega \in J_i$ which is decoupled from the controls associated with the block D_i. Since the system is assumed to be strongly accessible, this would imply that there is a one-form in J_i which is influenced by one of the other control blocks, and this contradicts our assumption that the outputs H_i are decoupled from the inputs D_j for $j \neq i$. Finally, it is straightforward to verify that

$$\sum_{j \neq i} J_j \subset I_i \subset M_i^\perp.$$

Therefore, we can conclude that if the system is strongly accessible and satisfies the noninteraction conditions, then it must satisfy

$$J_i \cap (\sum_{j \neq i} J_j) = \{dt\}.$$

□

If we remove the assumption that the system is strongly accessible, then things still work out in essentially the same way, but the condition that

$$J_i \cap (\sum_{j \neq i} J_j) = \{dt\}$$

is no longer necessary because two output blocks could both contain the same uncontrollable output.

It is insightful to look at the case when each output block involves only scalar output so that $dH_i = \text{span}\{dh^i, dt\}$. It will suffice to consider only the part of the system spanned by the differentials contained in J_i. Restricted to this subset, we have the matrix

$$\begin{bmatrix} I & G^i_{i+1}(\lambda_i, t) & F^i(\lambda_i, t) \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which is expressed relative to the ordered cobasis $d\hat{\lambda}, d\tilde{\lambda}, dt$. The codistribution $F^\perp$ is spanned by covectors of the form $\theta^j = \omega^j - (f \lrcorner \omega^j)dt$ where f is the drift vector, and $\omega^j \in F^\perp + \text{span}\{dt\}$. Each iteration of Algorithm 10.15 is of the form

$$I_{i(k+1)} = I_{ik} + L_F(F^\perp \cap I_{ik}).$$

For the first iteration, we must compute $I_{i0} \cap F^\perp$ This will have dimension 0 or 1 depending on whether or not $dh^i \in F^\perp + \text{span}\{dt\}$. If this condition is true, then one of the ω can be replaced by dh^i and we can compute that $I_{i1} = \{dh^i, dL_f h^i, dt\}$. If this condition is not true, then the iteration ends. Continuing in this fashion we find that at each step, we either add one differential to $I_{i(k+1)}$, or the iteration ends. Eventually, we find that

$$I_i = \text{span}\{dh^i, dL_f h^1, \ldots, dL_f^{r_i - 1} h^i, dt\},$$

and that

$$I_i \cap F^\perp = \text{span}\{dh^i - L_f H^i dt, dL_f h^1 - L_f^2 h^i dt, \ldots, dL_f^{r_i - 2} h^i - L^{r_i - 1} h^i dt\}.$$

Furthermore, using the basis described by 11.3 we find that $g_j \lrcorner L_f^{r_i - 1} h^i dt = \delta^i_j$. Therefore, the matrix

$$\begin{bmatrix} L_{g_1} L_f^{r_1 - 1} h^1 & \cdots & L_{g_m} L_f^{r_1 - 1} h^1 \\ L_{g_1} L_f^{r_2 - 1} h^2 & \cdots & L_{g_m} L_f^{r_2 - 1} h^2 \\ \vdots & \ddots & \vdots \\ L_{g_1} L_f^{r_p - 1} h^p & \cdots & L_{g_m} L_f^{r_p - 1} h^p \end{bmatrix}$$

has full rank. This is called the decoupling matrix. Its rank is invariant under feedback transformations, so the fact the outputs $h^1, \ldots, h^p$ can be rendered noninteractive implies that this matrix has full rank. Conversely, it is not difficult to show that if this matrix has full rank, then there exists a feedback transformation and a decomposition of the inputs which renders the system noninteractive. A control system which satisfies these conditions in an open neighborhood is said to have vector relative degree on this neighborhood. From the above discuss, it is apparent that when each output block consists of a scalar output, then the noninteracting control problem is solvable if and only if the system has vector relative degree on the region of interest.

Just as in the disturbance decoupling problem, even if there does not exist a static state feedback transformation which solves the noninteracting controls problem, there may exist a dynamic state feedback which does solve the problem. If so, it can be found by starting with the codistribution $dH_1 + dH_2 + \cdots + dH_p$, and applying Algorithm 10.3.1 to obtain the prolonged system.

Chapter 12

Conclusion

12.1 Contributions of this Dissertation

In closing, perhaps it is appropriate to outline the contributions which, in the author's opinion, this dissertation makes to the knowledge base of nonlinear control theory. While it is not widely used, the theory of exterior differential systems certainly is not new. Elements of the theory existed in the early 1800's, the first modern formulation is due to the work of Cartan around the turn of the century. Also, I am fairly sure that all of the results pertaining to nonlinear control theory can be found in other sources - though they may be stated somewhat differently. Therefore, the contribution of this thesis lies primarily in the connections which it establishes between the theory of exterior differential systems and the theory of nonlinear control systems. More specifically, the connection between Grassmann bundles and affine control systems, and the description of invariance within this framework. The disturbance decoupling and noninteracting control problems have been presented primarily to show how " standard " nonlinear controls problems can be effectively treated using this theory. Therefore, the theorems presented cover fairly standard results, but the proofs are in many cases quite different from what one would find in a standard treatment. Ultimately, the true contribution of any work must be measured by its readers. In this sense, it is too early to render any judgments. All I can say in this regards is that I have come away from this project with a deeper, more geometric understanding of nonlinear control systems, and I sincerely hope this presentation has been able to convey some of my insights and enthusiasm.

12.2 Future Directions

Although this dissertation has advocated the use of exterior differential systems theory in the study of nonlinear control theory and has in some cases presented comparisons between the vector field approach and the exterior differential systems approach to modeling nonlinear control systems, the intent has been more to supplement the vector field approach than to supplant it. In fact, if the exterior differential systems methodology gains in popularity, it will almost surely be due to the fact that it helps to bridge the gap between the vector field approach and other techniques for dealing with nonlinear systems such as optimal control theory and input-output analysis. Although the exterior differential systems approach is similar to the vector field approach in many ways, it also shares some common features with these other methodologies since it uses the state-time space. There has been some work done in this area - for example, the book by Griffiths [24]. However, this remains a wide-open area.

Another interesting area of investigation would be a study of the connections between exterior differential systems theory and classical mechanics. Problems in classical mechanics provided the initial impetus for many of the people who developed exterior differential systems theory, and it is closely associated with topics like contact transformations. Also, it is probably possible to give topics like virtual displacements and virtual work a more rigorous foundation using this theory. From a controls engineering standpoint, this type of investigation would also be useful since it would result in a common set of tools for analyzing mechanical hardware and controllers.

Up to this point, stability has not been mentioned at all in this dissertation. The reason for this omission is that stability is a relative property, i.e., a flow is only stable relative to something: an equilibrium point, another trajectory, a limit cycle, etc. However, it should be relatively straightforward to come up with a useful definition of stability on the state-time manifold and to discuss things like Lyapunov stability and bounded-input bound-output stability. A major part of this investigation would have to involve picking an appropriate metric to measure the distance between any two solution trajectories.

Finally, the Grassmann bundle description of a control system may provide a convenient method of describing uncertainty in system parameters, and it may even be possible to provide a of description of dynamic uncertainty within this framework. Again, a major part of the investigation would have to involve picking an appropriate metric to describe the magnitude of the error.

Bibliography

[1] R. B. Gardner and W. F. Shadwick. The GS algorithm for exact linearization to Brunovsky normal form. *IEEE Transactions on Automatic Control,* 37(2):224-230, 1992.

[2] R. M. Murray and S. S. Sastry. Nonholonomic motion planning: Steering using sinusoids. *IEEE Transactions on Automatic Control,* 38(5):700-716, 1993.

[3] D. Tilbury, R. M. Murray, and S. S. Sastry. Trajectory generation for the N-trailer problem using Goursat normal form. In *IEEE Control and Decision Conference,* 1993.

[4] A. Isidori. *Nonlinear Control Systems: An Introduction.* Springer-Verlag, 1989.

[5] J. R. Munkres. *Analysis on Manifolds.* Addison-Wesley, 1991.

[6] M. Spivak. *A Comprehensive Introduction to Differential Geometry.* Publish or Perish, 1979.

[7] K. Yang. *Exterior Differential Systems and Equivalence Problems.* Kluwer Academic Publishers, 1992.

[8] R. L. Bryant, S. S. Chern, R. B. Gardner, H. L. Goldschmidt, and P. A. Griffiths. *Exterior Differential Systems.* Mathematical Sciences Research Institute. Springer-Verlag, 1991.

[9] T. W. Hungerford. *Algebra.* Graduate Texts in Mathematics. Springer-Verlag, 1974.

[10] H. Flanders. *Differential Forms with Applications to the Physical Sciences.* Dover, 1989.

[11] H.J.C. Huijberts. *Dynamic Feedback in Nonlinear Synthesis Problems.* Amsterdam, Netherlands : Centrum voor Wiskunde en Informatica tract, 1994.

[12] D. Tilbury. *Exterior Differential Systems and Nonholonomic Motion Planning* Ph.D. thesis, University of California, Berkeley, 1994.

[13] M. Spivak. *Calculus on Manifolds.* Mathematics Monograph Series. Addison-Wesley, 1968.

[14] R. Murray. Applications and extensions of goursat normal form to control of nonlinear systems. In *Proceedings of the IEEE Control and Decision Conference*, 1993.

[15] D. Husemoller. *Fibre Bundles.* Graduate texts in mathematics; 20. 3rd edition, New York : Springer-Verlag, 1994.

[16] R. Abraham and J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications.* Applied Mathematical Sciences. Springer-Verlag, 1988.

[17] G. Pappas, J. C. Gerdes, and D. Niemann. Introduction to exterior differential systems. In *Advanced topics in adaptive and nonlinear control : Final projects EECS290B*, Electronics Research Laboratory memorandum UCB/ERL M95/8, University of California, Berkeley, 26 January, 1995.

[18] D. Tilbury and S. S. Sastry. *On Goursat Normal Forms, Prolongations, and Control Systems.* Electronics Research Laboratory memorandum UCB/ERL M94/16, University of California, Berkeley, 1994.

[19] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick. *Analysis, Manifolds, and Physics.* Elsevier, 1982.

[20] W. M. Sluis *Absolute Equivalence and its Application to Control Theory.* Ph.D. thesis, University of Waterloo, 1992.

[21] W. M. Sluis. A necessary condition for dynamic feedback linearization. *Systems & Control Letters*, 21(4): 277-83, 1993.

[22] W. Sluis and D. Tilbury. A bound on the number of integrators needed to linearize a control system. In *Proceedings of the 34th IEEE Conference on Decision and Control.* New Orleans, LA, USA, 13-15 Dec., 1995.

[23] H. Nijmeijer and A. van der Schaft. *Nonlinear Dynamical Control Systems.* Springer-Verlag, 1990.

[24] P. Griffiths. *Exterior Differential Systems and the Calculus of Variations.* Progress in mathematics, vol. 25. Birkhauser, 1982.

[25] R. J. Mihalek. *Projective Geometry and Algebraic Structures.* Academic Press, 1972.

Appendix A

Manifolds and the Tangent Bundle

This appendix is intended to provide a condensed treatment of basic topics in Differential Geometry. For a more detailed treatment, the reader may wish to consult one of the several books on the subject such as [6, 16, 5, 19].

A.1 Differentiable Manifolds

A manifold M of dimension n is a metric space which is locally homeomorphic to $\mathcal{R}^n$.

The simplest example of a manifold is $\mathcal{R}^n$ itself. Other examples are the circle S^1 and the sphere S^2. The circle S^1 is locally homeomorphic to $\mathcal{R}$ while the sphere is locally homeomorphic to $\mathcal{R}^2$. Therefore the circle is a one dimensional manifold while the sphere is a two dimensional manifold.

A subset N of manifold M which is itself a manifold is called a submanifold of M. Any open subset N of a manifold M is clearly a submanifold since if M is locally homeomorphic to $\mathcal{R}^n$ then so is N.

A function $f : \mathcal{R}^m \to \mathcal{R}^n$ is said to be smooth, or C^∞ at a point $p \in \mathcal{R}^m$ if its partial derivatives of all orders exist and are continuous at p. If this is true at every point of an open subset $V \subset \mathcal{R}^m$, then the function f is said to be C^∞ on V. In order to define a notion of smoothness for a mapping between two manifolds, we need to endow the manifolds with some additional structure.

A coordinate chart on a manifold M is a pair (U, x) where U is an open set of M and x is a homeomorphism of U on an open set of $\mathcal{R}^n$. The function x is also called a coordinate function and can also written as $(x^1, \ldots, x^n)$ where $x^i : M \longrightarrow \mathcal{R}$. If $p \in U$ then $x(p) = (x^1(p), \ldots, x^n(p))$ is called the set of local coordinates in the chart (U, x).

Two charts (U, x) and (V, y) with $U \cap V \neq \emptyset$, are called C^∞ compatible if the map

$$y \circ x^{-1} : x(U \cap V) \subset \mathcal{R}^n \longrightarrow y(U \cap V) \subset \mathcal{R}^n$$

is a C^∞ function. A C^∞ atlas on a manifold M is a collection of charts (U_α, x_α) with $\alpha \in A$ which cover the manifold and are pairwise C^∞ compatible. An atlas is called maximal if it is not contained in any other atlas. A differentiable or smooth manifold is a manifold with a maximal, C^∞ atlas.

Let $f : M \longrightarrow \mathcal{R}$ be any real-valued function defined on M. If (U, x) is a chart on M then the function

$$\hat{f} = f \circ x^{-1} : x(U) \subset \mathcal{R}^n \longrightarrow \mathcal{R}$$

is called the local representative of f in the chart (U, x). We therefore define the map f to be C^∞ or smooth if its local representative $\hat{f}$ is C^∞. Notice that if f is C^∞ in one chart, then it must be C^∞ in every chart since we required our charts to be C^∞ compatible and our atlas to be maximal. Similarly, if we have a map $f : M \longrightarrow N$, where M,N are differentiable manifolds, the local representation of f given a chart (U, x) of M and (V, y) of N is

$$\hat{f} = y \circ f \circ x^{-1}$$

which makes sense only if $f(U) \cap V \neq \emptyset$. Again, f is a C^∞ map if $\hat{f}$ is a C^∞ map, and this property of f is independent of the particular coordinate charts used to to construct $\hat{f}$.

Let $f : M \longrightarrow N$ be a map between two manifolds. The map f is called a diffeomorphism if both f and f^{-1} are smooth. In this case, manifolds M and N are called diffeomorphic.

Example: We have seen that $\mathcal{R}^n$ is an example of a trivial but important manifold. The differentiable structure on $\mathcal{R}^n$ consists of the chart $(\mathcal{R}^n, i)$ where i is the identity function on $\mathcal{R}^n$ as well as all other charts that are C^∞ compatible with it.

The sphere, S^2 can be given a differentiable structure as follows. Consider the charts (U_N, p_N) and (U_S, p_S) where U_N is the sphere minus the North pole, U_S is the sphere minus the South pole and p_N, p_S are the stereographic projections of the sphere to the plane from the North and South poles respectively. One can show that these charts are compatible. We can then extend our atlas to a maximal one by consider all other charts that are compatible with $(U_N, p_N), (U_S, p_S)$. ◇

A.2 The Tangent Bundle of a Smooth Manifold

Let p be a point on a manifold M. Let $C^\infty(p)$ denote the set of all smooth functions in a neighborhood of p. The set $C^\infty(p)$ is a vector space over $\mathcal{R}$ since the sum of two smooth functions and the scalar multiple of a smooth function are smooth function themselves.

A tangent vector X_p at $p \in M$ is an operator from $C^\infty(p)$ to $\mathcal{R}$ which satisfies for $f, g \in C^\infty(p)$ and $a, b \in \mathcal{R}$, the following properties,

1. Linearity $X_p(a \cdot f + b \cdot g) = a \cdot X_p(f) + b \cdot X_p(g)$
2. Derivation $X_p(f \cdot g) = f(p) \cdot X_p(g) + X_p(f) \cdot g(p)$

The set of all tangent vectors at $p \in M$ is called the tangent space of M at p and is denoted by T_pM.

The tangent space T_pM becomes a vector space over $\mathcal{R}$ if for tangent vectors X_p, Y_p and real numbers c_1, c_2 we define

$$(c_1 \cdot X_p + c_2 \cdot Y_p)(f) = c_1 \cdot X_p(f) + c_2 \cdot Y_p(f)$$

for any smooth function f in the neighborhood of p. The collection of all tangent spaces of the manifold,

$$TM = \bigcup_{p \in M} T_pM$$

is called the tangent bundle.

Example: Given the standard differentiable structure on $\mathcal{R}^n$, the standard tangent vectors of $\mathcal{R}^n$ at any point p are

$$\frac{\partial}{\partial r^1} \cdots \frac{\partial}{\partial r^n}$$

Thus given any smooth function $f(r^1, \ldots, r^n) : U \longrightarrow \mathcal{R}$ where U is a neighborhood of p, we have

$$\frac{\partial}{\partial r^i}(f) = \frac{\partial f}{\partial r^i}$$

for $i = 1, \ldots, n$. ◇

Now let M be a manifold and let (U, x) be a chart containing the point p. In this chart we can associate the following tangent vectors

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$$

defined by

$$\frac{\partial}{\partial x^i}(f) = \frac{\partial (f \circ x^{-1})}{\partial r^i}$$

for any smooth function $f \in C^\infty(p)$.

Theorem A.2.1 *Let M be an n dimensional manifold and let T_pM be the tangent space at $p \in M$. Then T_pM is an n-dimensional vector space and if (U, x) is a local chart around p then the tangent vectors*

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$$

form a basis for T_pM.

Proof: See Spivak [6] pages 107-108. □

From the above theorem we can see that if X_p is a tangent vector at p then

$$X_p = \sum_{i=1}^{n} a_i \frac{\partial}{\partial x^i}$$

where $a_1, \ldots, a_n$ are real numbers. From the above formula we can see that a tangent vector is an operator which simply takes the directional derivative of function in the direction of $[a_1, \ldots, a_n]$.

Now let M and N be smooth manifolds and $f : M \longrightarrow N$ be a smooth map. Let $p \in M$ and let $q = f(p) \in N$. We wish to transport tangent vectors from T_pM to T_qN using the map f. The natural way to do this is by defining a map $f_* : T_pM \longrightarrow T_qN$ by

$$(f_*(X_p))(g) = X_p(g \circ f)$$

for smooth functions g in the neighborhood of q. One can easily check that $f_*(X_p)$ is a linear operator and a derivation and thus a tangent vector. The map $f_* : T_pM \longrightarrow T_{f(p)}N$ is called the push forward map of f.

Theorem A.2.2 *The push forward map $f_* : T_pM \longrightarrow T_{f(p)}N$ is a linear map.*

Proof: Let X_p and Y_p be two tangent vectors in T_pM. Then

$$\begin{aligned} (f_*(X_p + Y_p))(g) &= (X_p + Y_p)(g \circ f) \\ &= X_p(g \circ f) + Y_p(g \circ f) \\ &= (f_*(X_p))(g) + (f_*(Y_p))(g) \end{aligned}$$

and also for real number c,

$$\begin{aligned} (f_*(c \cdot X_p))(g) &= (c \cdot X_p)(g \circ f) \\ &= c \cdot X_p(g \circ f) \\ &= c \cdot (f_*(X_p))(g) \end{aligned}$$

which completes the proof. □

Theorem A.2.3 *Let $f : M \longrightarrow N$ and $g : N \longrightarrow K$. Then*

$$(g \circ f)_* = g_* \circ f_*$$

Proof: See Spivak [6] page 101. □

Appendix B

Algebras and Ideals

We begin by introducing some algebraic structures which will be used in the development of the exterior algebra.

Definition B.0.1 *An* <u>*algebra*</u> *is a vector space* V *together with a multiplication operation* $\odot : V \times V \to V$ *which for every scalar* $\alpha \in \mathcal{R}$ *and* $a, b \in V$ *satisfies* $\alpha(a \odot b) = (\alpha a) \odot b = a \odot (\alpha b)$.

Definition B.0.2 *Given an algebra* $(V, \odot)$*, a subspace* $W \subset V$ *is called an algebraic ideal if* $x \in W, y \in V$ *implies that* $x \odot y, y \odot x \in W$

Note that if W is an ideal and $x, y \in W$ then $x + y \in W$ since W is a subspace.

Example: The set of all polynomials with real-valued coefficients, $\mathcal{R}[s]$, is a vector space over $\mathcal{R}$ with vector addition and scalar multiplication defined by

$$(P_1 + P_2)(s) = P_1(s) + P_2(s)$$

$$(\alpha \cdot P)(s) = \alpha \cdot P(s)$$

If we define multiplication by

$$(P_1 \cdot P_2)(s) = P_1(s) \cdot P_2(s)$$

then $\mathcal{R}[s]$ is also an algebra.

In $\mathcal{R}[s]$, the set of all polynomials with a zero at $s = -2$ is an algebraic ideal. This is true because for all $P_1(s), P_2(s) \in \mathcal{R}[s]$ which satisfy

$$P_1(-2) = P_2(-2) = 0$$

we have that

$$P_1(-2) + P_2(-2) = 0, \;\; \alpha \cdot P_1 = 0, \;\; P_1(-2) \cdot P_2(-2) = 0$$

so this set is a subspace of $\mathcal{R}[s]$ which is closed under multiplication. Furthermore for all $P(s), R(s) \in \mathcal{R}[s]$ with $R(-2) = 0$ we have that

$$P(-2) \cdot R(-2) = 0$$

which verifies that the set of all polynomials with a root at -2 is an ideal of $\mathcal{R}[s]$. $\diamond$

Definition B.0.3 *Let* $(V, \odot)$ *be an algebra. Let the set* $A := \{a_i \in V, 1 \leq i \leq K\}$ *be any finite collection of linearly independent elements in* V*. Let* S *be the set of all ideals containing* A

$$S := \{I \subset V | I \textit{ is an ideal and } A \subset I\}.$$

The ideal I_A generated by A is defined as

$$I_A = \bigcap_{I \in S} I$$

and is the minimal ideal in S containing A.

If $(V, \odot)$ is an algebra, and there exists an element $e \in V$ such that for all $x \in V, x \odot e = e \odot x = x$ then e is called a unity element and is unique. If $(V, \odot)$ is an algebra with a unity element, then the ideal generated by a finite set of elements can be represented in a simple form.

Theorem B.0.4 *Let $(V, \odot)$ be an algebra, $A := \{a_i \in V, 1 \leq i \leq K\}$ a finite collection of elements in V, and I_A the ideal generated by A. Then for each $x \in I_A$, there exist vectors $v_1, \ldots, v_K$ such that*

$$x = v_1 \odot a_1 + v_2 \odot a_2 + \ldots + v_K \odot a_K$$

Proof: See Hungerford [9] pages 123-124. □

Example: The polynomial $(s+2)$ generates an ideal in $\mathcal{R}[s]$ which is equal to the set of all polynomials with a zero at $s = -2$. We will denote this set as I_{-2}. In the previous example we verified that this set is an ideal, and the polynomial $(s+2)$ is clearly contained in I_{-2}. Therefore, in order to verify that I_{-2} is the ideal generated by $(s+2)$ we only need to show that any other ideal which contains $(s+2)$ also contains I_{-2}. Because a real root can always be factored, I_{-2} can be written as

$$I_{-2} := \{P(s) \in \mathcal{R}[s] \mid P(s) = R(s)(s+2) \, \forall \, R(s) \in \mathcal{R}[s]\} \tag{B.1}$$

If I is any other ideal containing $(s+2)$, then

$$(\forall \, R(s) \in \mathcal{R}[s]) \; R(s)(s+2) \in I$$

because of the definition of an ideal. Therefore, $I_{-2} \subset I$. Consequently, I_{-2} must be the ideal generated by $(s+2)$.

This result also follows directly from Theorem B.0.4, since $\mathcal{R}[s]$ has the constant polynomial **1** as a unity element. In order to see the importance of the unity element, suppose that we had taken as our algebra the set I_{-3} of all polynomials in $\mathcal{R}[s]$ with a root at -3. It is easy to verify that this is an algebra, and that the set $I_{-3,-2}$ of all polynomials with roots at -2 and -3 is an ideal. However,

$$I_{-3,-2} \neq \{P(s) \in \mathcal{R}[s] \mid P(s) = R(s)(s+2)(s+3) \, \forall \, R(s) \in I_{-3}\}$$

because the set on the right contains only polynomials with roots of order 2 and higher at -3. ◇

Example: The two polynomials $P_1(s) = (s+2)(s+4)$ and $P_2(s) = (s+2)(s+3)$ generate an ideal in $\mathcal{R}[s]$

$$I_{P_1,P_2} := \{P(s) \in \mathcal{R}[s] \mid P(s) = Q(s)P_1(s) + R(s)P_2(s), \; R(s), Q(s) \in \mathcal{R}[s]\}.$$

Although this ideal is generated by two linearly independent vectors, it is equivalent to the ideal generated by the single vector $P_3(s) = (s+2)$. To demonstrate this fact, let $P(s) \in I_{P_1,P_2}$. Then

$$P(s) = (Q(s)(s+3) + R(s)(s+4))(s+2).$$

which implies that $P(s) \in I_{-2}$.

Now suppose $P(s) \in I_{-2}$ Then

$$P(s) = Q(s)(s+2)$$

for some $Q(s) \in \mathcal{R}[s]$. From the coprime factorization property of polynomials it can be shown that there exists polynomials $N(s), M(s) \in \mathcal{R}[s]$ such that

$$1 = N(s)(s+3) + M(s)(s+4).$$

Using this identity we get that

$$P(s) = Q(s) \cdot 1 \cdot (s+2) = Q(s)N(s)(s+3)(s+2) + Q(s)M(s)(s+4)(s+2).$$

which implies that $P(s) \in I_{P_1,P_2}$. This example shows that if we are given an arbitrary set of generators, it may be possible to find a smaller set of generators which will generate the same ideal. $\diamond$

Definition B.0.4 *Let $(V, \odot)$ be an algebra, and $I \subset V$ an ideal. Two vectors $x, y \in V$ are said to be equivalent mod I if and only if $x - y \in I$. This equivalence is denoted*

$$x \equiv y \ mod \ I$$

From the definition above we can see that

$$x \equiv y \ mod \ I$$

if and only if

$$x - y \in I$$

which simply means that

$$x - y = \sum_{i=1}^{K} \theta_K \odot \alpha_K$$

for some $\theta_K \in V$. It is customary to abuse notation and denote this as

$$x \equiv y \ mod \ \alpha_1, \ldots, \alpha_K$$

where the mod operation is implicitly performed over the ideal generated by $\alpha_1, \ldots, \alpha_K$.